GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:27:49 ; Search time 2545.54 Seconds
(without alignments)
11432.162 Million cell updates/sec

Title: US-09-579-383-2
Perfect score: 1764
Sequence: 1 atgaattttaaaatatcaat.....tggatttaccattacaataa 1764

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB	ID	Description
1	1764	100.0	2500	3	AF064079	AF064079 Plasmodiu
2	1047.2	59.4	1764	12	AF216703	AF216703 Synthetic
3	937.2	53.1	2562	3	PBE305256	AJ305256 Plasmodiu
c 4	196	11.1	245802	2	AC006279	AC006279 Plasmodiu
5	182	10.3	1325	3	AF072442	AF072442 Plasmodiu
c 6	104.4	5.9	14867	3	AE001398	AE001398 Plasmodiu
7	104.2	5.9	1137	3	AF172445	AF172445 Plasmodiu
8	102.8	5.8	104992	2	AC005504	AC005504 Plasmodiu
9	102.8	5.8	169546	2	AC004157	AC004157 Plasmodiu
10	102.6	5.8	67970	3	PFMAL1P3	AL031746 Plasmodiu
c 11	97.6	5.5	213530	2	AC090493	AC090493 Mus muscu
c 12	93.8	5.3	183584	9	AC012492	AC012492 Homo sapi
c 13	93.6	5.3	86827	3	PFMAL3P5	AL034556 Plasmodiu
14	93.6	5.3	199551	2	AC006281	AC006281 Plasmodiu
c 15	92	5.2	47577	3	AF396436	AF396436 Tetrahyme
16	92	5.2	170143	2	AC048384	AC048384 Homo sapi
c 17	92	5.2	178273	2	AC005308	AC005308 Plasmodiu
c 18	91.8	5.2	143331	9	AC091214	AC091214 Homo sapi
19	91.8	5.2	178719	2	AL137250	AL137250 Homo sapi
20	91.4	5.2	4601	3	DMU11584	U11584 Drosophila
21	91.4	5.2	19517	3	DMU37541	U37541 Drosophila
c 22	91.4	5.2	326924	2	AC093082	AC093082 Homo sapi
c 23	90.4	5.1	236120	14	AF063866	AF063866 Melanoplu
c 24	89.2	5.1	180063	2	AC074281	AC074281 Homo sapi
25	88.4	5.0	326924	2	AC093082	AC093082 Homo sapi
26	88.2	5.0	156933	2	AC026839	AC026839 Homo sapi
c 27	88	5.0	54345	3	AC084152	AC084152 Caenorhab
28	88	5.0	86827	3	PFMAL3P5	AL034556 Plasmodiu
29	88	5.0	218078	2	AC068138	AC068138 Homo sapi
c 30	87.8	5.0	67970	3	PFMAL1P3	AL031746 Plasmodiu
31	87.2	4.9	14001	3	PFCOMPIRB	X95276 P.falciparu
32	87	4.9	14867	3	AE001398	AE001398 Plasmodiu
c 33	86.8	4.9	142239	2	AL390835	AL390835 Homo sapi
c 34	86.6	4.9	840	8	CNS018QK	AL110675 Botrytis
c 35	86.6	4.9	170143	2	AC048384	AC048384 Homo sapi
c 36	86.2	4.9	120029	2	HSJ282H10	AL132672 Homo sapi
c 37	86.2	4.9	172805	9	AC066608	AC066608 Homo sapi
38	85.8	4.9	180063	2	AC074281	AC074281 Homo sapi
39	85.6	4.9	106763	9	AP002091	AP002091 Homo sapi
40	85.6	4.9	161286	2	AC025120	AC025120 Homo sapi
41	85.4	4.8	110000	2	PFMAL4P1_1	Continuation (2 of
42	85.2	4.8	170928	9	AC034148	AC034148 Homo sapi
c 43	84.6	4.8	175223	9	AC004617	AC004617 Homo sapi
44	84.6	4.8	217930	2	AC026471	AC026471 Homo sapi
45	84.4	4.8	825	8	SCCOX1	X00418 Yeast cytoc

ALIGNMENTS

RESULT 1
AF064079
LOCUS AF064079 2500 bp mRNA INV 10-APR-2000
DEFINITION Plasmodium gallinaceum endochitinase precursor, mRNA, complete cds.
ACCESSION AF064079
VERSION AF064079.1 GI:7530423
KEYWORDS .
SOURCE Plasmodium gallinaceum.
ORGANISM Plasmodium gallinaceum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 2500)
AUTHORS Vinetz,J.M., Valenzuela,J.G., Specht,C.A., Aravind,L., Langer,R.C., Ribeiro,J.M. and Kaslow,D.C.
TITLE Chitinases of the avian malaria parasite Plasmodium gallinaceum, a class of enzymes necessary for parasite invasion of the mosquito midgut
JOURNAL J. Biol. Chem. 275 (14), 10331-10341 (2000)
MEDLINE 20209408
REFERENCE 2 (bases 1 to 2500)
AUTHORS Vinetz,J.M., Valenzuela,J. and Kaslow,D.C.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1998) Pathology, University of Texas Medical
            Branch at Galveston, Keiller 2.138, 301 University Blvd, Galveston,
            TX 77555-0609, USA
FEATURES             Location/Qualifiers
     source          1. .2500
                     /organism="Plasmodium gallinaceum"
                     /db_xref="taxon:5849"
                     /dev_stage="ookinete"
     CDS             276. .2039
                     /EC_number="3.2.1.14"
                     /codon_start=1
                     /product="endochitinase precursor"
                     /protein_id="AAF63208.1"
                     /db_xref="GI:7530424"
                     /translation="MNFKISIFLIIVSILYSANSRTLKGKNNINNSLGIIRENKNKTH
                     QTEIHESFSHLKSNNSNFVEYGSYCGDGCNSRITKNNKNINKNDRKSPRQILEEYKKR
                     KQGIIAGYYGSWNSQGDRAKHMIDSNPMVSILYIAFARINMLYDVSRPFNGRQRFLLR
                     KHGLEYETYGMMLNEIRRIRKVRPDVIILLSLGGETYMIDIEKEIDYVDKILKLVNDF
                     DLDGVDIDWEPHGKFYNLNELNFSNYYIKLINLLRKTIPEEKLISISGSSNAALSCVS
                     GVASFCKDEESPYNTKFLSEQIETNKELHRAAAMLSAGTFINIFNTAKEKIDLVFIQT
                     YNLETTNPDIMVDMYLSHLYFGLKYNITIILGFSLEHNRGGFSPENKELLELVGKTIH
                     DKNQNNNRADGIGIWHLFMKEQLPTGSFDVDIFLTNIWKHLNPEVQTPKDLTITENPE
                     DCSTIDEYVPGLVIPTIGIYYKHNDAIWKTRSYSIHAPGVDRYEWDLVKVCYEKICDG
                     KAAHYYNTDYKESSIIIWKGEPYLIKWWQQGPPEGQALESYTKLDASKCPGIEEWNKK
                     YPHKPLEVEEQYEQEVDLPLQ"
BASE COUNT      993 a    283 c    365 g    858 t      1 others
ORIGIN
```

Query Match 100.0%; Score 1764; DB 3; Length 2500;
Best Local Similarity 100.0%; Pred. No. 2.4e-242;
Matches 1764; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgaattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattcc 60
Db     276 ATGAATTTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTCTGCAAATTCC 335

Qy      61 agaaccttgaaaggaaaaaatataataaataattcattgggaataatacgggaaaataaa 120
Db     336 AGAACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGAAAATAAA 395

Qy     121 aataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagt 180
Db     396 AATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGT 455

Qy     181 aattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaat 240
Db     456 AATTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAAT 515

Qy     241 aataaaaatataaataaaaatgatagaaaatcaccaagacaaatttagaggagtataaa 300
Db     516 AATAAAAATATAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAA 575

Qy     301 aaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgataga 360
Db     576 AAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGA 635

Qy     361 gcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgc 420
Db     636 GCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGC 695

Qy     421 attaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaaga 480
Db     696 ATTAATATGTTATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGA 755

Qy     481 aaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcaga 540
Db     756 AAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGA 815

Qy     541 aaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagat 600
Db     816 AAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGAT 875

Qy     601 atagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgattta 660
Db     876 ATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTA 935

Qy     661 gatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 720
Db     936 GATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAAT 995

Qy     721 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 780
Db     996 TTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAG 1055

Qy     781 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 840
Db    1056 TTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCT 1115

Qy     841 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 900
Db    1116 TTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 1175

Qy     901 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960
Db    1176 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTT 1235

Qy     961 aatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1020
Db    1236 AATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACTACA 1295

Qy    1021 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1080
Db    1296 AATCCAGATATAATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATATAAC 1355

Qy    1081 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1140
Db    1356 ATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAAT 1415

Qy    1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200
Db    1416 AAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAATAGG 1475

Qy    1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1260
Db    1476 GCAGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAACTGGATCATTC 1535

Qy    1261 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1320
Db    1536 GATGTAGATATTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAAGTACAAACTCCA 1595

Qy    1321 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1380
Db    1596 AAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAGCACAATAGATGAATATGTTCCA 1655

Qy    1381 ggactcgttattccaaccatagggatatattacaaacacaatgatgctatatggaaaact 1440
Db    1656 GGACTCGTTATTCCAACCATAGGGATATATTACAAACACAATGATGCTATATGGAAAACT 1715

Qy    1441 agatcttattcaattcatgcacctggtgtagacagatatgaatgggacttggtcaaagta 1500
Db    1716 AGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAAAGTA 1775

Qy    1501 tgctatgaaaaaatatgcgatgggaaagcagcccattattataacactgactataaagaa 1560
Db    1776 TGCTATGAAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGACTATAAAGAA 1835

Qy    1561 agctctattattatatggaaaggggaaccatatttaattaaatggtggcaacaaggacct 1620
Db    1836 AGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCT 1895

Qy    1621 ccggaaggtcaggcactagagtcatacacaaaactagatgcatccaaatgtccagggata 1680
Db    1896 CCGGAAGGTCAGGCACTAGAGTCATACACAAAACTAGATGCATCCAAATGTCCAGGGATA 1955
```

```
Qy   1681 gaagaatggaataaaaaatatccacataaaccactagaagtagaggaacaatatgaacaa 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1956 GAAGAATGGAATAAAAAATATCCACATAAACCACTAGAAGTAGAGGAACAATATGAACAA 2015

Qy   1741 gaagtggatttaccattacaataa 1764
          |||||||||||||||||||||||
Db   2016 GAAGTGGATTTACCATTACAATAA 2039


RESULT   2
AF216703
LOCUS       AF216703     1764 bp    DNA             SYN       26-JUN-2000
DEFINITION  Synthetic construct chitinase CHT1 (Cht1) gene, complete cds.
ACCESSION   AF216703.
VERSION     AF216703.1  GI:8705223
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1764)
  AUTHORS   Vinetz,J.M.
  TITLE     Chitinases of the avian malaria parasite, Plasmodium gallinaceum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1764)
  AUTHORS   Vinetz,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1999) WHO Center for Tropical Disease, University
            of Texas Medical Branch, 301 University Blvd., Galveston, TX
            77555-0609, USA
FEATURES             Location/Qualifiers
     source          1. .1764
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Plasmodium gallinaceum gene engineered for
                     bacterial codon usage"
     gene            1. .1764
                     /gene="Cht1"
     CDS             1. .1764
                     /gene="Cht1"
                     /codon_start=1
                     /transl_table=11
                     /product="chitinase CHT1"
                     /protein_id="AAF78777.1"
                     /db_xref="GI:8705224"
                     /translation="MNFKISIFLIIVSILYSANSRTLKGKNNINNSLGIIRENKNKTH
                     QTEIHESFSHLKSNNSNFVEYGSYCGDGCNSRITKNNKNINKNDRKSPRQILEEYKKR
                     KQGIIAGYYGSWNSQGDRAKHMIDSNPMVSILYIAFARINMLYDVSRPFNGRQRFLLR
                     KHGLEYETYGMMLNEIRRIRKVRPDVIILLSLGGETYMIDIEKEIDYVDKILKLVNDF
                     DLDGVDIDWEPHGKFYNLNELNFSNYYIKLINLLRKTIPEEKLISISGSSNAALSCVS
                     GVASFCKDEESPYNTKFLSEQIETNKELHRAAAMLSAGTFINIFNTAKEKIDLVFIQT
                     YNLETTNPDIMVDMYLSHLYFGLKYNITIILGFSLEHNRGGFSPENKELLELVGKTIH
                     DKNQNNNRADGIGIWHLFMKEQLPTGSFDVDIFLTNIWKHLNPEVQTPKDLTITENPE
                     DCSTIDEYVPGLVIPTIGIYYKHNDAIWKTRSYSIHAPGVDRYEWDLVKVCYEKICDG
                     KAAHYYNTDYKESSIIIWKGEPYLIKWWQQGPPEGQALESYTKLDASKCPGIEEWNKK
                     YPHKPLEVEEQYEQEVDLPLQ"
BASE COUNT      545 a    422 c    386 g    411 t
ORIGIN


  Query Match             59.4%;  Score 1047.2;  DB 12;  Length 1764;
  Best Local Similarity   74.6%;  Pred. No. 2.5e-140;
  Matches 1316;  Conservative    0;  Mismatches  448;  Indels    0;  Gaps    0;

Qy      1 atgaattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattcc 60
          |||||||| ||||| || || ||  | || || ||   |||| |||||   || || |||
Db      1 ATGAATTTCAAAATCTCCATCTTCCTGATCATCGTTAGCATCCTGTATAGCGCCAACTCC 60

Qy     61 agaaccttgaaaggaaaaaatatatataaataattcattgggaataatacgggaaaataaa 120
            | ||| ||||||| || || || || || || ||  |||| || || || ||||| |||
Db     61 CGTACCCTGAAAGGTAAGAACAACATCAACAACTCCCTGGGTATTATCCGCGAAAACAAA 120

Qy    121 aataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagt 180
          || ||||| |||||||| ||||| |||||||| ||||| ||  | ||||| ||||||||

Db    121 AACAAAACCCATCAAACCGAAATCCATGAGTCCTTTTCTCACTTAAAATCCAATAATAGC 180

Qy    181 aattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaat 240
          ||||| || || || ||     ||||| || ||||| ||||||     | || || |||||
Db    181 AATTTCGTTGAGTACGGTAGCTATTGTGGCGATGGCTGTAACAGCCGTATCACGAAAAAC 240

Qy    241 aataaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaa 300
          || ||||| || |||||||| ||  | |||   ||  | || ||| | || || || |||
Db    241 AACAAAAACATCAATAAAAACGACCGCAAAAGCCCGCGCCAGATTCTCGAAGAATACAAA 300

Qy    301 aaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgataga 360
          ||| | ||||| || ||||| || ||||||||||| |||||||| || || ||  |
Db    301 AAACGCAAACAGGGCATTATTGCCGGCTACTATGGTTCCTGGAACAGCCAGGGCGACCGT 360

Qy    361 gcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgc 420
          || ||||| ||||| || || || || |||||||| ||  | ||||| || || || ||  |
Db    361 GCCAAACACATGATCGACTCTAATCCGATGGTGTCCATCCTGTATATCGCGTTCGCCCGT 420

Qy    421 attaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaaga 480
          ||||||||| | ||||| || ||| | || ||||| ||  | ||  | || ||  |  |
Db    421 ATTAATATGCTGTATGACGTTTCTCGTCCGTTTAACGGTCGCCAGCGTTTTCTGCTGCGT 480

Qy    481 aaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcaga 540
          || || ||| | ||||| |||||||||||| |||||||| ||||||||| | || ||| |
Db    481 AAGCATGGCCTGGAATACGAAACCTATGGCATGATGCTGAATGAAATTCGTCGCATCCGC 540

Qy    541 aaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagat 600
          ||||||||| || || |||| ||||| ||| |||||| || |||||||||||||| ||
Db    541 AAAGTACGCCCGGACGTAATCATTCTGTTAAGCTTAGGCGGCGAAACCTATATGATCGAC 600

Qy    601 atagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgattta 660
          || |||||||||| || ||||| |||||||  |||| || || || |||||||| |
Db    601 ATCGAAAAAGAAATCGACTATGTCGATAAAATTCTGAAACTGGTGAACGATTTTGATCTG 660

Qy    661 gatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 720
          |||||||| || || |||||||||| ||||| ||||| |||||  | |||||| | ||
Db    661 GATGGTGTTGACATCGACTGGGAACCGCATGGCAAGTTCTACAATCTGAATGAACTGAAC 720

Qy    721 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 780
          ||    ||||| ||||| ||| | |||||| || |  | ||||| || ||||| |||||
Db    721 TTCAGCAATTACTATATCAAACTGATTAACCTGCTGCGCAAAACCATCCCGGAGGAAAAA 780

Qy    781 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 840
            | || || || ||||||       || || ||| |    ||||||   || || || |||
Db    781 CTGATCTCTATCTCTGGTAGCAGCAACGCAGCACTGAGCTGCGTTAGCGGCGTCGCTTCT 840

Qy    841 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 900
          ||||| |||||||||||||| || ||||||||||||||  ||   |||||||| |||||
Db    841 TTCTGCAAAGATGAAGAATCCCCGTATAACACTAAATTCCTGAGCGAACAAATCGAAACT 900

Qy    901 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960
          || |||||  | ||| | || || || ||| |    || || || || ||||| |||||
Db    901 AACAAAGAGCTGCATCGTGCTGCCGCAATGCTGAGCGCCGGTACGTTCATTAACATTTTC 960

Qy    961 aatacagcaaaggagaaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1020
          || || || || || ||||| || || |||||||| ||||| || ||| | || || ||
Db    961 AACACTGCGAAAGAAAAAATCGACCTGGTATTTATCCAAACCTATAATCTGGAGACCACC 1020

Qy   1021 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1080
          || || ||||| ||||| |||||||| |   |||  | ||||| ||  | ||||| |||
Db   1021 AACCCGGATATTATGGTCGATATGTATCTGAGCCACCTGTATTTCGGCCTGAAATACAAC 1080

Qy   1081 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1140
          || || |||||  | |||||||  | ||||| ||  | || || || || || || ||
Db   1081 ATTACGATCATCCTGGGTTTTTCTCTGGAACACAATCGTGGCGGTTTCAGCCCGGAGAAC 1140

Qy   1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200
          |||||| |  | |||||||| || ||||| || || || || || || ||||||||  |
Db   1141 AAAGAACTGCTGGAATTGGTTGGCAAAACTATCCACGACAAGAACCAGAATAATAACCGT 1200

Qy   1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1260
          || |||||||| || || |||||| | || ||||| |||||| | || |||||    ||
Db   1201 GCTGATGGTATCGGTATTTGGCATCTGTTCATGAAGGAACAACTGCCGACTGGTAGCTTT 1260
```

```
Qy   1261 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1320
          || || |||||||||||| || || |||||||||||| | || || ||||| ||||| ||
Db   1261 GACGTGGATATTTTTCTGACGAACATTTGGAAACATCTGAACCCGGAAGTGCAAACCCCG 1320

Qy   1321 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1380
          |||||| | || || || |||||||| |||||||| ||| || || || ||||| ||
Db   1321 AAAGACTTAACCATCACCGAAAACCCGGAAGACTGCTCCACGATCGACGAGTATGTACCG 1380

Qy   1381 ggactcgttattccaaccatagggatatattacaaacacaatgatgctatatggaaaact 1440
          || || || ||||| ||||| || || || || |||||||| ||||| || ||||||||
Db   1381 GGTCTGGTGATTCCGACCATCGGCATCTACTATAAACACAACGATGCGATTTGGAAAACC 1440

Qy   1441 agatcttattcaattcatgcacctggtgtagacagatatgaatgggacttggtcaaagta 1500
          | || ||    || |||||||| || || ||  | || || |||||  |||| || ||
Db   1441 CGTTCCTACAGCATCCATGCACCGGGCGTCGATCGCTACGAGTGGGATCTGGTGAAGGTT 1500

Qy   1501 tgctatgaaaaaatatgcgatgggaaagcagcccattattataacactgactataaagaa 1560
          |||||||| ||||| || ||||| ||||| || |||||||||||| || ||||||||||||
Db   1501 TGCTATGAGAAAATTTGTGATGGCAAAGCTGCGCATTATTATAATACCGACTATAAAGAA 1560

Qy   1561 agctctattattatatggaaaggggaaccatatttaattaaatggtggcaacaaggacct 1620
          |||||||||||| || ||||| || || || || ||||| |||||||||||| ||||| ||
Db   1561 AGCTCTATTATCATTTGGAAGGGTGAGCCGTACTTAATCAAATGGTGGCAGCAAGGCCCG 1620

Qy   1621 ccggaaggtcaggcactagagtcatacacaaaactagatgcatccaaatgtccagggata 1680
          |||||||||| || || |||   ||||| ||||| ||||| ||||| ||||| || ||
Db   1621 CCGGAAGGTCAAGCCCTGGAGAGCTACACCAAACTGGATGCTTCCAAGTGTCCGGGCATT 1680

Qy   1681 gaagaatggaataaaaaatatccacataaaccactagaagtagaggaacaatatgaacaa 1740
          |||||||||||| || || ||||| || || || || ||||| || ||||| || |||||
Db   1681 GAAGAATGGAACAAGAAGTATCCGCACAAGCCGCTGGAAGTTGAAGAACAGTACGAACAG 1740

Qy   1741 gaagtggatttaccattacaataa 1764
          ||||||||  | ||  | ||||||
Db   1741 GAAGTGGACCTGCCGCTGCAATAA 1764
```

```
RESULT   3
PBE305256
LOCUS       PBE305256    2562 bp    mRNA            INV       30-MAY-2001
DEFINITION  Plasmodium berghei mRNA for chitinase (cht1 gene).
ACCESSION   AJ305256
VERSION     AJ305256.1  GI:14275848
KEYWORDS    chitinase; ChT1 gene.
SOURCE      Plasmodium berghei.
  ORGANISM  Plasmodium berghei
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 2562)
  AUTHORS   Dessens,J.T., Mendoza,J., Claudianos,C., Vinetz,J.M., Khater,E.,
            Hassard,S., Ranawaka,G.R. and Sinden,R.E.
  TITLE     Knockout of the rodent malaria parasite chitinase pbcht1 reduces
            infectivity to mosquitoes
  JOURNAL   Infect. Immun. 69 (6), 4041-4047 (2001)
  MEDLINE   21246711
REFERENCE   2  (bases 1 to 2562)
  AUTHORS   Dessens,J.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Dessens J.T., Dept. of Biology, Imperial
            College of Science, Technology and Medicine, SAFB, Imperial College
            Road, London, SW7 2AZ, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2562
                     /organism="Plasmodium berghei"
                     /strain="ANKA"
                     /isolate="2.34"
                     /db_xref="taxon:5821"
     sig_peptide     174. .227
                     /gene="cht1"
     CDS             174. .2120
                     /gene="cht1"
                     /function="mosquito infection"
                     /codon_start=1
                     /product="chitinase"
                     /protein_id="CAC40151.1"
                     /db_xref="GI:14275849"
                     /translation="MNIYVSLFLIIFSTINNAYSVGLKGGTTINHPINISHTEKQYKS
                     LSHVDALCGDSKNYVCETKYICKNGSKGEYCSEGGSGNGAENGNGNGVENGNGVENGN
                     GAENGNGAENGNGNGVENGNGNGAENGNGNGAENGNGNGVENGNGNGNHERKSPREIL
                     EEYKKRKQGIIAGYYGSWNSQGASGKQMTHSNPHVSIIYIAFARINMNYDASKPYNGL
                     QKFLQKKHGLEFETYGMMANEIKRIRKARPDVIIILSLGGETYMIDITKEIDYIPQIV
                     NMVNNFDLDGVDIDWEPNGKFYHLNELAFSNYYIKLIDLIRSNISKDKLISISGSSNA
                     ALSCVSVWETFCKDEETPYNTQYLSEQMIVNPELYKASTMLSTGTFVNVFNTAKDKID
                     LVFIQTYNLETSNPDIMLDMYLSHLYFGKKYDITVFLGFSIEQNRGGFSPDSKNLVEL
                     VSKAIHDNNHEFNRADGVGIWHLFMKEQMPNGAYDLDGFIDNTWTHLNPGIKSPTDIV
                     SIPNPDDCSTIEEYVQGVVIPTVGTYAKHNDAIWVTRSYSTRAPGIDRYEWNLVKVCY
                     EKACNNEAAHYYNVDYPTGTKIIWKGEVFSIKQWHGGPPEGPYLESYEKLDEYSCPGL
                     AEWNIKHPHKPIEVEAPYEQELDG"
     gene            174. .2120
                     /gene="cht1"
     mat_peptide     228. .2117
                     /gene="cht1"
                     /product="chitinase"
BASE COUNT     1001 a    323 c    416 g    822 t
ORIGIN
```

```
  Query Match           53.1%;  Score 937.2;  DB 3;  Length 2562;
  Best Local Similarity  72.7%;  Pred. No. 1.1e-124;
  Matches 1225;  Conservative    0;  Mismatches  458;  Indels    3;  Gaps    1;

Qy      79 aataatataaataattcattgggaataatacgggaaaataaaaataaaactcatcaaacg 138
           ||| | || | ||| | || || | | |||| ||| | |
Db     441 AATGGTGTAGAAAATGGAAATGGTGTAGAAAATGGAAATGGTGCAGAAAATGGAAATGGT 500

Qy     139 gaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgtagaatatgga 198
           | | | |||    |       | | ||     ||| | ||| || |||| |||||
Db     501 GCAGAAAATGGAAATGGGAATGGTGTAGAAAATGGAAATGGAAATGGTGCAGAAAATGGA 560

Qy     199 tct---tattgcggagatgggtgtaactctagaattacaaaaaatataaaaatataaat 255
             |    || | | |||    | ||    |    | | |||||   ||   ||
Db     561 AATGGAAATGGTGCAGAAAATGGAAATGGGAATGGTGTAGAAAATGGAAATGGTAATGGA 620

Qy     256 aaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaaaggaaacaaggt 315
           ||  |||| |||||||| |||||  |||||||||| || ||||||||||| || ||||||
Db     621 AACCATGAAAGAAAATCTCCAAGGGAAATTTTAGAAGAATATAAAAAAAGAAAGCAAGGT 680

Qy     316 attatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacatatgatt 375
           || || |||||||| ||||| |||||||| || ||||| | ||| |  |||| |||| |
Db     681 ATAATTGCAGGATATTATGGATCATGGAATAGCCAAGGAGCTAGCGGGAAACAAATGACT 740

Qy     376 gattcaaacccaatggtgtcaatttttatatattgcatttgctcgcattaatatgttatat 435
           |||| ||||||   || || ||  | ||||| || || || || || |||||   |||
Db     741 CATTCCAACCCACACGTATCTATAATTTATATAGCTTTCGCCCGTATAAATATGAACTAT 800

Qy     436 gatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggcttagaa 495
           |||| |||| ||||| |||||||   |||| |||  ||  || ||| || ||  |||||
Db     801 GATGCTTCTAAACCATATAATGGACTTCAAAAATTTTTACAAAAAAAGCATGGACTAGAA 860

Qy     496 tatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagat 555
           | |||||| |||||||||||| |||||||||| |||||| |||||||  || ||||||
Db     861 TTTGAAACATATGGTATGATGGCTAATGAAATTAAACGTATAAGAAAAGCCCGCCCAGAT 920

Qy     556 gtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaaagaaatt 615
           ||||||||  |||||||| |||||  |||||||| || ||||||||||||  ||||||||||
Db     921 GTAATTATAATTTTATCATTAGGAGGAGAAACATACATGATAGATATAACAAAAGAAATT 980

Qy     616 gattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgtagatatt 675
           |||||| |   | ||||| | ||  | |||||  |||||||||||||||||||||||||
Db     981 GATTATATTCCTCAAATAGTAAACATGGTTAACAATTTTGATTTAGATGGTGTAGATATT 1040

Qy     676 gactgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaaattattat 735
           || ||||||||| |||| ||||||||| | ||||||||| |    ||||||||||||||
Db    1041 GATTGGGAACCAAATGGAAAGTTTTACCATTTAAATGAACTCGCCTTTTCAAATTATTAC 1100

Qy     736 attaaattaattaacttgttaagaaaaactattccggaagaaaagttaatttcaatttct 795
```

```
Db   1101 ATTAAATTGATTGATTTAATAAGAAGCAACATTTCTAAAGATAAGCTAATTTCCATATCT 1160

Qy    796 ggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttctgtaaagatgaa 855
Db   1161 GGATCATCAAATGCAGCATTATCATGTGTATCAGTATGGGAGACATTTTGTAAAGATGAA 1220

Qy    856 gaatctccatataacactaaatttttgtctgaacaaatagaaacaaataaagaattacat 915
Db   1221 GAAACGCCATATAATACACAGTATTTATCTGAACAAATGATCGTAAACCCTGAATTATAC 1280

Qy    916 agggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaaggag 975
Db   1281 AAAGCATCAACCATGTTATCTACAGGAACATTTGTTAATGTTTTTAATACAGCAAAGGAT 1340

Qy    976 aaaatagatcttgtatttattcaaacatacaatttagaaactacaaatccagatataatg 1035
Db   1341 AAAATTGATCTTGTTTTTATTCAAACATATAATTTAGAAACATCAAACCCAGATATCATG 1400

Qy   1036 gtagatatgtacttatcccatttatattttggtttaaaatataacatcacaatcatatta 1095
Db   1401 CTTGATATGTATTTATCTCATTTGTATTTTGGAAAAAAATATGATATAACTGTTTTCTTA 1460

Qy   1096 ggtttttcattagaacataacagaggtggatttagtcccgaaaataaagaattattagaa 1155
Db   1461 GGTTTTTCAATAGAACAAAATAGAGGTGGATTTAGTCCAGATAGCAAAAACTTAGTAGAA 1520

Qy   1156 ttggtaggaaaaacaatacatgataaaaatcaaaataataatagggcagatggtataggg 1215
Db   1521 TTAGTATCGAAAGCTATACATGACAATAATCACGAATTCAACAGAGCCGATGGTGTAGGA 1580

Qy   1216 atatggcatttatttatgaaagaacaattaccaactggatcattcgatgtagatattttt 1275
Db   1581 ATTTGGCATTTATTTATGAAAGAACAAATGCCTAATGGAGCATATGATCTAGATGGTTTT 1640

Qy   1276 cttacaaatatttggaaacatttaaatcctgaagtacaaactccaaaagaccttactata 1335
Db   1641 ATTGATAATACATGGACCCATTTAAATCCAGGAATAAAATCCCCAACAGATATAGTTTCA 1700

Qy   1336 actgaaaaccctgaagactgtagcacaatagatgaatatgttccaggactcgttattcca 1395
Db   1701 ATCCCTAACCCTGACGATTGTAGTACTATAGAAGAATATGTACAAGGAGTCGTTATCCCA 1760

Qy   1396 accatagggatatattacaaacacaatgatgctatatggaaaactagatcttattcaatt 1455
Db   1761 ACTGTAGGAACATATGCTAAACACAATGATGCTATATGGGTAACAAGGTCCTATTCAACA 1820

Qy   1456 catgcacctggtgtagacagatatgaatgggacttggtcaaagtatgctatgaaaaaata 1515
Db   1821 CGGGCACCTGGTATCGACAGATATGAATGGAACTTAGTAAAAGTATGTTATGAAAAAGCA 1880

Qy   1516 tgcgatgggaaagcagcccattattataacactgactataaagaaagctctattattata 1575
Db   1881 TGTAATAATGAAGCTGCACATTATTATAACGTTGACTACCCAACAGGAACAAAGATCATT 1940

Qy   1576 tggaaaggggaaccatatttaattaaatggtggcaacaaggacctccggaaggtcaggca 1635
Db   1941 TGGAAAGGAGAAGTATTCTCTATTAAACAATGGCATGGAGGTCCACCGGAAGGCCCTTAT 2000

Qy   1636 ctagagtcatacacaaaactagatgcatccaaatgtccagggatagaagaatggaataaa 1695
Db   2001 CTTGAATCATATGAAAAATTAGACGAATACAGCTGCCCAGGATTAGCAGAATGGAATATA 2060

Qy   1696 aaatatccacataaaccactagaagtagaggaacaatatgaacaagaagtggatttacca 1755
Db   2061 AAACACCCACATAAACCAATTGAAGTTGAAGCTCCATATGAGCAAGAACTTGATGGATAA 2120

Qy   1756 ttacaa 1761
Db   2121 ATACAA 2126

RESULT   4
```

```
AC006279/c
LOCUS       AC006279   245802 bp    DNA             HTG       12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 5 unordered pieces.
ACCESSION   AC006279
VERSION     AC006279.8  GI:9797723
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 245802)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245802)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810451.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   147568: contig of 147568 bp in length
            *  147569   147768: gap of unknown length
            *  147769   217470: contig of 69702 bp in length
            *  217471   217670: gap of unknown length
            *  217671   219372: contig of 1702 bp in length
            *  219373   219572: gap of unknown length
            *  219573   221333: contig of 1761 bp in length
            *  221334   221533: gap of unknown length
            *  221534   245802: contig of 24269 bp in length.
FEATURES             Location/Qualifiers
     source          1. .245802
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC614"
                     /clone="3D7"
BASE COUNT    99718 a  25978 c  21493 g  97813 t    800 others
ORIGIN


  Query Match           11.1%;  Score 196;  DB 2;  Length 245802;
  Best Local Similarity  53.1%;  Pred. No. 1.7e-19;
  Matches  543;  Conservative    0;  Mismatches  455;  Indels   24;  Gaps    5;

Qy     265 agaaaatcaccaagacaaattttagaggagtataaaaaaaggaaacaaggtattatagca 324
Db   45336 AGAAAAAACCCGAGAGAAATTATAAAAACATTTAAAGAGAGTGGGAAAGGAATTATTCAA 45277

Qy     325 ggatactatggttcatggaacagtcaaggtgatagagcaaaacatatgattgattcaaac 384
Db   45276 GGTTATTATCCATCATGGGTTTCTTATAATCATA---------ACTTGAAAGATTTAAAT 45226

Qy     385 ccaatggtgtcaattttatatattgcatttgctcgcattaatatgttatatgatgtatct 444
Db   45225 CCTAATTTAAATGTTGTACATATGTCTTTTGCTAAGATGGATTTATCCTATGACAGTATA 45166

Qy     445 agaccatttaatggaagacaaagattcctattaagaaaacacggcttagaatatgaaacc 504
Db   45165 GAGAGTATTGTTGGTTCTCCTTTACTTTTCAAGTCATTAATAGGTTTAGAATATATAGGG 45106

Qy     505 tatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagatgtaattatt 564
Db   45105 TTAAACGAATATTTTAATGATGCCATGAATTTAAGAAAAGCTCGTCCAGATATTATTATG 45046
```

```
Qy    565 cttttatccttaggtggagaaacctatatgatagatata---gaaaaagaaattgattat 621
Db  45045 CTTTTATCTCTTGGAGGAGAAACATATCATCCAAGTTCATTTGATTCTGCTTTAAATGCT 44986

Qy    622 gtggataaaatattgaagcttgttaatgattttgatttagatggtgtagatattgactgg 681
Db  44985 GTTGAGAAAATTGCAAATTTAGTTGATGAATTAGGATTTGATGGAATCGATGTTGATTAT 44926

Qy    682 gaaccacatgggaagttttacaacttaaatgaatttaaatttttcaaattattatattaaa 741
Db  44925 GAGCCTAATGGATCTTTTGATGGTTTGAACGATAAAGAGAAAGCTGATTTTTTCGTACAA 44866

Qy    742 ttaattaacttgttaagaaaaactattccggaagaaaagttaatttcaatttctggttca 801
Db  44865 TATGTAACAAAATTGAGGGAATATATGTGTGATGATAAATTAATTTCTATATCTCAATCA 44806

Qy    802 tcaaatgctgcattatcatgcgtttcaggagttgcatct------ttctgtaaagatgaa 855
Db  44805 TCTAACGGAGCTTTAAGTTGTATAGGTTTTAATGATCCTAAGAAAATATGTATGGATGAT 44746

Qy    856 gaatctccatataacactaaatttttgtctgaacaaatagaaacaaataaagaattacat 915
Db  44745 GAAGCACCATATAATTCAAAATATTTCAATAAAC---CTGATGTAAAAAAGGAGCTATTA 44689

Qy    916 agggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaaggag 975
Db  44688 AGAGCAGCGCAAATGGCTTCAGCTGGAGGAGCAATATATTTAATGAATAATTTAAAGGAT 44629

Qy    976 aaaatagatcttgtatttattcaaacatacaatt---tagaaactacaaatccagatata 1032
Db  44628 ATGATTGATATGGTGTTTGTACAAACCTTTAATTATACAAATTCTACAGATTCAACAGTT 44569

Qy   1033 atggtagatatgtacttatcccatttatattttggtttaaaatataacatcacaatcata 1092
Db  44568 ATGAAAGAATTATATGATTCTTATGCTTATTATGGAAAAAAATATGATTATGTTATTATA 44509

Qy   1093 ttaggtttttcattagaacataacagaggtggatttagtcccgaaaataaagaattatta 1152
Db  44508 ATGGGATTTACATTAATGTTTCCTTCAACCCCTTTTAATCCGAATGATAAAATGCTTGTA 44449

Qy   1153 gaattggtaggaaaaacaatacatgataaaaatcaaaataataatagggcagatggtata 1212
Db  44448 AAATCCATTGGTGATTTTGTAAAGACTGAAAATAAATTAAATAAAAGAGCAGACGGGTTT 44389

Qy   1213 gggatatggcatttatttatgaaagaacaattaccaactggatcattcgatgtagatatt 1272
Db  44388 GGATTATGGTCTTTATCTTCTGATAATGCTGCACATAATGAACAATTAGCAATTGAATAT 44329

Qy   1273 tt 1274
Db  44328 TT 44327
```

```
RESULT   5
AF072442
LOCUS       AF072442     1325 bp    DNA             INV       10-APR-2000
DEFINITION  Plasmodium falciparum chitinase gene, partial cds.
ACCESSION   AF072442
VERSION     AF072442.1  GI:7530426
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1325)
  AUTHORS   Vinetz,J.M., Valenzuela,J.G., Specht,C.A., Aravind,L., Langer,R.C.,
            Ribeiro,J.M. and Kaslow,D.C.
  TITLE     Chitinases of the avian malaria parasite Plasmodium gallinaceum, a
            class of enzymes necessary for parasite invasion of the mosquito
            midgut
  JOURNAL   J. Biol. Chem. 275 (14), 10331-10341 (2000)
  MEDLINE   20209408
REFERENCE   2  (bases 1 to 1325)
  AUTHORS   Vinetz,J.M. and Kaslow,D.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-1998) Laboratory of Parasitic Diseases, National
            Institute of Allergy and Infectious Disease, 9000 Rockville Pike,
            Bethesda, MD 20892-0425, USA
FEATURES             Location/Qualifiers
     source          1. .1325
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="14"
                     /clone="3D7"
                     /note="TIGR clone PNADI77"
     mRNA            <1. .>971
                     /product="chitinase"
     CDS             <1. .>971
                     /note="includes the putative catalytic domain, but not the
                     chitin binding domain"
                     /codon_start=1
                     /product="chitinase"
                     /protein_id="AAF63209.1"
                     /db_xref="GI:7530427"
                     /translation="KGIIQGYYPSWVSYNHNLKDLNPNLNVVHMSFAKMDLSYDSIES
                     IVGSPLLFKSLIGLEYIGLNEYFNDAMNLRKARPDIIMLLSLGGETYHPSSFDSALNA
                     VEKIANLVDELGFDGIDVDYEPNGSFDGLNDKEKADFFVQYVTKLREYMCDDKLISIS
                     QSSNGALSCIGFNDPKKICMDDEAPYNSKYFNKPDVKKELLRAAQMASAGGAIYLMNN
                     LKDMIDMVFVQTFNYTNSTDSTVMKELYDSYAYYGKKYDYVIIMGFTLMFPSTPFNPN
                     DKMLVKSIGDFVKTENKLNKRADGFGLWSLSSDNAAHNEQLAIEYFV"
BASE COUNT      474 a    130 c    185 g    536 t
ORIGIN
```

```
  Query Match             10.3%;  Score 182;  DB 3;  Length 1325;
  Best Local Similarity   53.1%;  Pred. No. 3.4e-17;
  Matches  486;  Conservative    0;  Mismatches  415;  Indels   15;  Gaps    4;

Qy    371 tgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatgt 430
Db     53 TGAAAGATTTAAATCCTAATTTAAATGTTGTACATATGTCTTTTGCTAAGATGGATTTAT 112

Qy    431 tatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggct 490
Db    113 CCTATGACAGTATAGAGAGTATTGTTGGTTCTCCTTTACTTTTCAAGTCATTAATAGGTT 172

Qy    491 tagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtc 550
Db    173 TAGAATATATAGGGTTAAACGAATATTTTAATGATGCCATGAATTTAAGAAAAGCTCGTC 232

Qy    551 cagatgtaattattcttttatccttaggtggagaaacctatatgatagat---atagaaa 607
Db    233 CAGATATTATTATGCTTTTATCTCTTGGAGGAGAAACATATCATCCAAGTTCATTTGATT 292

Qy    608 aagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggtg 667
Db    293 CTGCTTTAAATGCTGTTGAGAAAATTGCAAATTTAGTTGATGAATTAGGATTTGATGGAA 352

Qy    668 tagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaa 727
Db    353 TCGATGTTGATTATGAGCCTAATGGATCTTTTGATGGTTTGAACGATAAAGAGAAAGCTG 412

Qy    728 attattatattaaattaattaacttgttaagaaaaactattccggaagaaaagttaattt 787
Db    413 ATTTTTTCGTACAATATGTAACAAAATTGAGGGAATATATGTGTGATGATAAATTAATTT 472

Qy    788 caatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct------t 841
Db    473 CTATATCTCAATCATCTAACGGAGCTTTAAGTTGTATAGGTTTTAATGATCCTAAGAAAA 532

Qy    842 tctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaa 901
Db    533 TATGTATGGATGATGAAGCACCATATAATTCAAAATATTTCAATAAACCTGATGTAAAAA 592

Qy    902 ataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatatttta 961
```

```
Db     593 AGGAGCTATTA---AGAGCAGCGCAAATGGCTTCAGCTGGAGGAGCAATATATTTAATGA 649

Qy     962 atacagcaaaggagaaaatagatcttgtatttattcaaacatacaatt---tagaaacta 1018
           |||    |||||| |  || ||| | || ||| | ||||| |  ||||    | |  |||
Db     650 ATAATTTAAAGGATATGATTGATATGGTGTTTGTACAAACCTTTAATTATACAAATTCTA 709

Qy    1019 caaatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatata 1078
           || || ||    | |||  |||  | ||    ||  ||   |||| |||   |||||||
Db     710 CAGATTCAACAGTTATGAAAGAATTATATGATTCTTATGCTTATTATGGAAAAAAATATG 769

Qy    1079 acatcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaa 1138
           |       || ||| | || ||| |||||     |     |      |||| |||  |
Db     770 ATTATGTTATTATAATGGGATTTACATTAATGTTTCCTTCAACCCCTTTTAATCCGAATG 829

Qy    1139 ataaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataata 1198
           |||||    |  || |||   | ||  |     || |   | ||||| ||   ||||| |
Db     830 ATAAAATGCTTGTAAAATCCATTGGTGATTTTGTAAAGACTGAAAATAAATTAAATAAAA 889

Qy    1199 gggcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcat 1258
           | ||||| ||  | ||  ||||| |||||  |    |  |      ||  | || |  ||
Db     890 GAGCAGACGGGTTTGGATTATGGTCTTTATCTTCTGATAATGCTGCACATAATGAACAAT 949

Qy    1259 tcgatgtagatatttt 1274
           | |   | ||   |||
Db     950 TAGCAATTGAATATTT 965


RESULT   6
AE001398/c
LOCUS       AE001398    14867 bp    DNA             INV       06-NOV-1998
DEFINITION  Plasmodium falciparum chromosome 2, section 35 of 73 of the
            complete sequence.
ACCESSION   AE001398 AE001362
VERSION     AE001398.1  GI:3845197
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 14867)
  AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
            Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
            Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M.,
            Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
            Smith,H.O., Fraser,C.M., Hoffman,S.L. et,al.
  TITLE     Chromosome 2 sequence of the human malaria parasite Plasmodium
            falciparum
  JOURNAL   Science 282 (5391), 1126-1132 (1998)
  MEDLINE   99021743
  REMARK    Erratum:[[published erratum appears in Science 1998 Dec
            4;282(5395):1827]]
REFERENCE   2  (bases 1 to 14867)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
            Medical Center Drive, Rockville, MD 20814, USA
FEATURES             Location/Qualifiers
     source          1. .14867
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="2"
     gene            complement(1570. .2424)
                     /gene="PFB0490c"
     CDS             complement(1570. .2424)
                     /gene="PFB0490c"
                     /note="predicted by GlimmerM"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC71887.1"
                     /db_xref="GI:3845198"
                     /translation="MKEKNEKIMDYLSCPLDDVVDREKKSGKNSLLKSSSTKKSDYKK
                     SSIFSKKRDSHKKGSSFRGRRSGFINRKSGSFKKPYYNNRLINKNYNNYKGRNFHNGR
                     DNFKGRTGSFGSRVFDNRKGSFKKRFISNRNKSSVKSYRGNGSNKMGRKSFNKAPTSR
                     TVVTKRLNNYKTVSAPVKKFNNLNISLYRKNRTFALNTKRSKPVGTIKSSVPRKRIKK
                     GLKKGSLKSKTRKSTSGSKFKPLNKYFLSKIKIVTSLNKIPSPLKEQKNTEVNLPESL
                     NNATTKKN"
     gene            11241. .14606
                     /gene="PFB0495w"
     CDS             11241. .14606
                     /gene="PFB0495w"
                     /note="predicted by GlimmerM"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC71888.1"
                     /db_xref="GI:3845199"
                     /translation="MFALKKNTVREGFVNICFSYLKKLYLKSNFVTVNLNYETNNEKR
                     TNKKIYKKSKAQSLFDKGLNIHDKLILFKNLPKYKCAKYECISAKEVYKYLLDEYKKC
                     FNYISLCDIIQSVKIFDELDKTFTDYNFYIEVKNIDKNVLNKINEIYFKNKDITFHRR
                     EILGKICNKIMSYIHEMNGNELIHFLIYFFRWNKNDKNLILFYNYYFNYVFDHMYLFN
                     HEIYKLLFIFNKYLNNNSNIPFNKNLIQEMEFNLYYFREIKNEKNYIIKMNKKEIYKK
                     CFAKFHENVDHIDNEKILNILRLYVDNSILDIDINNKMLCNLNNNLINENIEYISKLL
                     NFYCTLIKKGKYDNDMTIYKLKEVIKATHHILCDKTKNLETFCSDIDYSTLLNSLNNK
                     FILNKIIDKNFILFYECLLKILLNIKFVNFQSLCISLISLKNIYYNILRNNVYIVNNV
                     LFNDIMKFSLYLCNIFLGKRIKTENENAVLIIHNNDQTNYSNKENIKDIIIQKRIKEY
                     IFYKMENYKDFHFKLKDSDLLSIKLLSNTFVKINEVYNSYDFYLLFNNISCILYNFLV
                     NRNSVKKYKDTYIYILNDLSFVYKYIKNNDRTKKKNFFLLSSSMKELICKNILSVSN
                     RYIKHLHEEDNFDQKDQYVCSLTFLNNLFFDKIIHFHYIYNLWCHVYKTYNYFKCNKL
                     INEDIISLLLLTCSKFQYFIENNSNDRYCRKELIHLKYNIIDDLIKNYLNTYKSISID
                     NISKIFISLSNSKYTCEVNENLLLESLQSEFEKVTKTSKKGGIHMMDNNLLDNNNSCE
                     KYEHRYIEYKKENLFINLNKIIECLIKLNIFLYLKKKKTYLYLYKQSLCPINLKENIL
                     KKILYIANNLYMYEMYGYVCEMLERVLSSHKEQNLFSYNYNKNVEHKMFDKILCHISE
                     DDYIEMSNTMYVLFYDYLKNINSERQSNILRNNSTNDRFIDEIKEKKYKLNNNTLIKH
                     NNVKLNYEKSNNSNGNISNILKDDKNKNHNNVEMDLIDNKNENKKIQEKGQNGENCEN
                     CKDVLVNDIINIFGFLKMEKKKFLFFQLYMYLCNITKFKRRYVSSSSLFHMDVFKIIK
                     DMNLKYLCLENYKIKNEECAFLYTIDIVLFKER"
BASE COUNT     6284 a   1019 c   1106 g   6458 t
ORIGIN


  Query Match             5.9%;  Score 104.4;  DB 3;  Length 14867;
  Best Local Similarity   44.6%;  Pred. No. 2.7e-06;
  Matches  635;  Conservative    0;  Mismatches  771;  Indels   18;  Gaps    5;

Qy       5 attttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaa 64
             || || |||||| ||| ||   || ||    ||  ||| |||   ||    |||||||||
Db    9803 ATGTTCAAATATTAATTTTACAAAGTAGTAAATTAATCATGTTACCTTATTATTCCAGAA 9744

Qy      65 ccttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaata 124
             || ||    ||||  |  |||| | |  |||     || ||     || || | |  |
Db    9743 AATTTAATTAGAAAACAATAATAAGTTAGACATAAAATATTATGTAAAAATATTACACGA 9684

Qy     125 aaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaatt 184
             | |   ||||   | |||   || |   | | |    |   || ||  |||  | || ||
Db    9683 ATATATATCATTTGTAAAATTATAAAATTATATTTATTAAAAATTCATATATCATTATTT 9624

Qy     185 ttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaatataata 244
             |  ||  | | |    ||       |   | |||       |       | | || |
Db    9623 CAATTCAAAGTTGCTAAATTTTATATATTATTTATAAT------AAATATTATATTATTT 9570

Qy     245 aaaatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaa 304
             || |||||||| ||| ||    |||   |  || | ||| | || |  |   |||| ||
Db    9569 AATATATAAATTAAATTGTATTAAATAAAAAAAAAAAAAATATATATTATGTTAAATAAT 9510

Qy     305 ggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaa 364
             | |  |  |||  |   |   |                   |   |  |  | | | |  ||
Db    9509 ATACATTATATATATTTTAATTTTTAGTATTAATTATTTAATATATATATTAAAAAAGAA 9450

Qy     365 aacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatta 424
             | || |  || ||  | |   |  |||| ||| | | ||||  |||   |  | ||
Db    9449 TATATTTTTTTTATAAATATAAATGTGTGTAAATATAAAATATAAGATTAAATAAAAATA 9390

Qy     425 atatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaac 484
             | || | | |  || | | || |  | |  ||   | | |   |||   | |||  | |
Db    9389 A-ATTTAAAAATATATTTTTAAATTAATATATAATAAATATTTATTAAAAATAATTATAT 9331

Qy     485 acggctta--gaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaa 542
               | |  ||||   ||   ||  || |     | | ||||   |  ||| | | |
```

```
Db    9330 TAAAATAAATAAATAATTAATTAATTAAATAAATAAAATTAAAATATTAAATATTAAATA 9271

Qy     543 agtacgtccagatgtaattattctttttatccttaggtggagaaacctatatgatagatat 602

Db    9270 AATTTAATATAATAAATTAAATTAAAAAATAATATATAAATAATATAAAATAAATAATTT 9211

Qy     603 agaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttaga 662

Db    9210 AATTAAAATAATAAATAAACTAAAATAATAAACAAATAATATAAAATAAATAAATTAAAA 9151

Qy     663 tggt--gtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 720

Db    9150 TAATAATTAAATAATATAAAATAAATAATTTAATTAAAATATAAAATAAATAAATTTAAA 9091

Qy     721 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 780

Db    9090 TAATAAATAAATAATATAAAAATAAATTAATATATTAAATAAATTAATTAAAATAATTAA 9031

Qy     781 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 840

Db    9030 ATAAATAAAA-------TAAAATAAATTAATATATTAAATAAATAAATTTAAATAATAAA 8978

Qy     841 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 900

Db    8977 TAAATAAAATAAAAATAAATAATATATTAAATAAATAAATAAAATAAAAATAAATTAATA 8918

Qy     901 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960

Db    8917 TATTAAATAAATTAATTAAAATAATTAAATAAATAAAATAAAAATAAATAAATAAATAAA 8858

Qy     961 aatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1020

Db    8857 TAAAATAAAAATAAATAAATAAAATAAAAATAAATTAATATATTAAATAAATTAATTAAA 8798

Qy    1021 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1080

Db    8797 ATAATAAATAAATAATATAAAATAAAATAATTAAATTAAAATAAGTAAAAATTAATTAAA 8738

Qy    1081 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1140

Db    8737 ATAAAATAAATAATTATGTGTAAATTATAATGTATATGTTTAAATTTTTATATACTAATT 8678

Qy    1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200

Db    8677 AAATTATTAATATATATAATATTATATCATAATTATTTTAAAAATAATATAATAAATATT 8618

Qy    1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1260

Db    8617 ATAATTAATGAATTAAAGCAATATTTATTTTATAAAAATAAATTAATAATTATATTAACT 8558

Qy    1261 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1320

Db    8557 TTAATAAATACATTATTTATTTATATTGGAAATAATTATTATTATATAAATATATATTAA 8498

Qy    1321 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1380

Db    8497 AATAAATGAAATAATATTATAATTTTTATTAATTAATATAATTTAATTTAATAATATAAA 8438

Qy    1381 ggactcgttattccaaccatagggatatattacaaacacaatga 1424

Db    8437 TAAAATAATAATATATTAATACAAAGAAACAAAATATATTATTA 8394


RESULT   7
AF172445
LOCUS       AF172445     1137 bp    DNA             INV       12-DEC-1999
DEFINITION  Plasmodium falciparum chitinase gene, complete cds.
ACCESSION   AF172445
VERSION     AF172445.1  GI:6561403
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Vinetz,J.M., Dave,S.K., Specht,C.A., Brameld,K.A., Xu,B.,
            Hayward,R. and Fidock,D.A.
  TITLE     The chitinase PfCHT1 from the human malaria parasite Plasmodium
            falciparum lacks proenzyme and chitin-binding domains and displays
            unique substrate preferences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (24), 14061-14066 (1999)
  MEDLINE   20040676
REFERENCE   2  (bases 1 to 1137)
  AUTHORS   Vinetz,J.M. and Dave,S.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-1999) Center for Tropical Diseases, University of
            Texas Medical Branch, 301 University Blvd., Galveston, TX
            77555-0609, USA
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="Plasmodium falciparum"
                     /isolate="K1"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /country="Thailand"
     mRNA            <1. .>1137
                     /product="chitinase"
     CDS             1. .1137
                     /codon_start=1
                     /product="chitinase"
                     /protein_id="AAF16902.1"
                     /db_xref="GI:6561404"
                     /translation="MNFTVKYSFLVICLLCCLLSTYVSVIEGHRARPGESRKNPREII
                     KTFKESGKGIIQGYYPSWVSYNHNLKDLNPNLNVVHMSFAKMDLSYDSIESIVGSPLL
                     FKSLIGLEYIGLNEYFNDAMNLRKARPDIIMLLSLGGETYHPSSFDSALNAVEKIANL
                     VDELGFDGIDVDYEPNGSFDGLNDKEKADFFVQYVTKLREYMCDDKLISISQSSNGAL
                     SCIGFNDPKKICMDDEAPYNSKYFNKPDVKKELLRAAQMASAGGAIYLMNNLKDMIDM
                     VFVQTFNYTNSTDSTVMKELYDSYAYYGKKYDYVIIMGFTLMFPSTPFNPNDKMLVKS
                     IGDFVKTENKLNKRADGFGLWSLSSDNAAHNEQLAIEYFVESLH"
BASE COUNT      321 a    263 c    257 g    296 t
ORIGIN


  Query Match             5.9%;  Score 104.2;  DB 3;  Length 1137;
  Best Local Similarity   48.0%;  Pred. No. 4e-06;
  Matches  468;  Conservative    0;  Mismatches  483;  Indels   24;  Gaps    5;

Qy     266 gaaaatcaccaagacaaattttagaggagtataaaaaaaggaaacaaggtattatagcag 325

Db     110 GCAAAAACCCGCGTGAAATTATCAAAACGTTTAAAGAATCCGGTAAAGGTATCATCCAGG 169

Qy     326 gatactatggttcatggaacagtcaaggtgatagagcaaaacatatgattgattcaaacc 385

Db     170 GTTACTATCCGTCCTGGGTGAGCTATAATCACA---------ACCTGAAGGACCTGAACC 220

Qy     386 caatggtgtcaattttatatattgcatttgctcgcattaatatgttatatgatgtatcta 445

Db     221 CGAACCTGAACGTGGTTCACATGAGCTTTGCTAAAATGGATTTAAGCTACGACTCCATTG 280

Qy     446 gaccatttaatggaagacaaagattcctattaagaaaacacggcttagaatatgaaacct 505

Db     281 AATCCATCGTAGGTAGCCCGCTGCTGTTTAAAAGCCTGATTGGCCTGGAGTATATTGGTC 340

Qy     506 atggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagatgtaattattc 565

Db     341 TGAATGAGTACTTCAATGATGCCATGAATCTGCGCAAAGCTCGTCCGGACATTATTATGC 400

Qy     566 ttttatccttaggtggagaaacct---atatgatagatatagaaaaagaaattgattatg 622

Db     401 TGCTGAGCCTGGGCGGTGAGACCTACCATCCGTCCTCCTTCGATTCTGCACTCAACGCGG 460

Qy     623 tggataaaatattgaagcttgttaatgattttgatttagatggtgtagatattgactggg 682

Db     461 TTGAAAAAATCGCAAATCTGGTGGATGAACTGGGCTTCGATGGTATTGATGTAGATTACG 520

Qy     683 aaccacatgggaagttttacaacttaaatgaattaaatttttcaaattattatattaaat 742
```

```
Db    521 AACCGAACGGCAGCTTTGATGGTCTGAATGATAAAGAAAAAGCGGACTTCTTTGTACAAT 580

Qy    743 taattaacttgttaagaaaaactattccggaagaaaagttaatttcaatttctggttcat 802
          | | |    |  |  ||   ||     || || ||  | ||    ||       || |
Db    581 ACGTGACCAAACTGCGCGAATACATGTGTGATGATAAACTGATCAGCATCAGCCAGTCCT 640

Qy    803 caaatgctgcattatcatgcgtttcaggagttgcatctt------tctgtaaagatgaag 856
          | ||||  ||  |   ||| |         |   |        ||||||  ||||| |
Db    641 CTAATGGCGCTCTGAGCTGCATCGGTTTCAACGACCCGAAAAAAATCTGTATGGATGACG 700

Qy    857 aatctccatataacactaaatttttgtctgaacaaatagaaacaaataaagaattacata 916
          || |||| |||||||  |||| |||     |  ||   ||    || ||||||  |
Db    701 AAGCTCCGTATAACAGCAAATATTT---CAACAAACCGGACGTTAAGAAAGAACTGTTAC 757

Qy    917 gggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaaggaga 976
          | |||||   ||||  ||| || ||   |   ||  |  |  | || |     || || |
Db    758 GCGCAGCCCAGATGGCATCTGCGGGTGGTGCCATCTACCTGATGAACAACCTGAAAGATA 817

Qy    977 aaatagatcttgtatttattcaaacatacaatt---tagaaactacaaatccagatataa 1033
          || ||  | || ||| | || || | ||| |      |  ||||  || |      | |
Db    818 TGATTGACATGGTGTTTGTGCAGACGTTCAACTACACGAACTCTACCGATTCTACCGTTA 877

Qy   1034 tggtagatatgtacttatcccatttatattttggtttaaaatataacatcacaatcatat 1093
          ||  |||  | |||   ||| |   ||| | ||||  ||| ||  |   |   |||||
Db    878 TGAAAGAATTATACGACTCCTACGCATACTATGGTAAAAAGTACGATTACGTGATCATCA 937

Qy   1094 taggtttttcattagaacataacagaggtggatttagtcccgaaaataaagaattattag 1153
          | || ||  | |                    || |  ||  |  |||||    |  ||
Db    938 TGGGCTTCACCCTGATGTTCCCGTCCACGCCGTTCAACCCGAACGATAAAATGCTGGTAA 997

Qy   1154 aattggtaggaaaaacaatacatgataaaaatcaaaataataatagggcagatggtatag 1213
          |||  | ||  |     || |   | ||||  ||   |||||  | ||||||||| | |
Db    998 AATCTATCGGCGATTTCGTAAAAACTGAAAACAAACTGAATAAACGCGCAGATGGTTTCG 1057

Qy   1214 ggatatggcatttat 1228
          |  | |||  | | |
Db   1058 GCCTGTGGTCTCTGT 1072


RESULT   8
AC005504
LOCUS       AC005504   104992 bp    DNA             HTG       01-APR-1999
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC005504
VERSION     AC005504.3  GI:4558584
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 104992)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
            and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 104992)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Apr 2, 1999 this sequence version replaced gi:4337172.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    58642: contig of 58642 bp in length
            *   58643    58842: gap of unknown length
            *   58843    91011: contig of 32169 bp in length
            *   91012    91211: gap of unknown length
            *   91212   104992: contig of 13781 bp in length.
FEATURES             Location/Qualifiers
     source          1. .104992
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT    44286 a   9326 c   9564 g  41411 t    405 others
ORIGIN


  Query Match             5.8%;  Score 102.8;  DB 2;  Length 104992;
  Best Local Similarity   45.5%;  Pred. No. 3.5e-06;
  Matches  662;  Conservative    0;  Mismatches  757;  Indels   35;  Gaps    7;

Qy      9 taaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaacctt 68
          ||| | ||  | |  | | | | || | | |  || ||  ||     ||| |    |  ||
Db  72725 TAATAAATAGAAAAATATTAATACAATTTAAATATTAAATAAATAAAAATATTATAATTT 72784

Qy     69 gaaaggaaaaaatatataaataattcattgggaataatacgggaaaata--aaaataaa 126
            ||  || |||  ||  | |||| | | |    |  |||    ||| ||  | ||| ||
Db  72785 ATAAATAATAAAATATTAATATAAATTAATTAATAATATATAATAAATTAATATAATTAA 72844

Qy    127 actcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaatttt 186
          | | |  | |    |  | | |    |  | | || ||||    ||  | | ||||  |
Db  72845 ATTTAAAATATAAATTAATAAAAAAATAATACTAATATTAATATAAAATAAAATAATAAT 72904

Qy    187 gtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaa 246
            | ||  |  |  | |      | |    ||    || || |  |  |||||||||
Db  72905 AAATAAATTTTAATTAAAATTAAATAATAAAATATTAATATAAATTAATTAAATAATAAT 72964

Qy    247 aatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaagg 306
          |  |||||| ||    |||  ||   |  ||   || |  |  |  |   | ||| ||
Db  72965 ATAATAAATTAATTAAATAATAATATAATAAATTAATTAATTAACAATATTTAAATAATT 73024

Qy    307 aaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaa 366
          ||| |   || ||||     |     |  ||  |  ||     |   | ||  |  | ||
Db  73025 AAATATAATAATATATATTAAACAATTAATTATTATAAATTAAAGAATTATTAATAATAA 73084

Qy    367 catatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaat 426
           |||| | | | | |||   ||||   |   |     ||        |  |    |||
Db  73085 TATATAAATTAATTAAATAAAATGAATTATTTAAATAATTAAAACAATAATATATATAAA 73144

Qy    427 atgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacac 486
           |  |     || ||   |    | | ||   || |   | |||  || |   |   |
Db  73145 TTAATTATATATTTAGTAAATAAAATAAAATTAATAATTAAATTAATAATTTATAGTTAT 73204

Qy    487 ggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagta 546
              | || |  |||  |||  |   ||   |||  |||||| |   ||    |||  |
Db. 73205 AAAAAATAAAAATAAATATATAATTAAATATATAAATAAATTAAATAAATATATAAATAA 73264

Qy    547 cgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatatagaa 606
           |  | || ||   |    || |     |  |  | | |  || || | | |||||  |
Db  73265 ATTAAATATATATACA---ATTAAATTAAATATATATATAATTAAATTAAATATATATAT 73321

Qy    607 aaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggt 666
          ||  ||||| | ||| |  ||||   | | ||  |  | ||| |  |  ||| | |
Db  73322 AATTAAATTAAATATATATATAATTAAATTAAATTAATAAATAAAATAAATTAATAAATA 73381

Qy    667 gtagat---attgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
            | ||   ||| | |    |     | ||    |  |  |  ||  || | || | | |
Db  73382 AAATATATAATTAAATTAAATAAATATATAATTAAATTAATATATAAATAAATTAAATAT 73441

Qy    724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
           | || |||| |||| ||||  ||| ||  ||| | | | |||     |  |||   ||
Db  73442 ATATATAATTAAATTATATTATATAAATTAATAATATATAATATAATAAATAAAATAATA 73501

Qy    784 atttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttc 843
          |||  || |  |  |  || | ||  |  || | ||    ||  |  | |   || ||
Db  73502 ATT--AAATAATAATATATAATATAATAAATAAACAATAATTAAATTAATAATATATTAT 73559

Qy    844 tgtaaagatgaagaatctccatataacactaaatttttgtctgaacaa-----atagaaa 898
            || || |  | |    |||  | | ||| |  ||   | || ||       |  ||
```

```
Db  73560 AAATAATATAATAATTAAATATAATATAATAATTAATTAAATTAATAATATATTAAATAA 73619

Qy    899 caaataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattt 958
          | |  || || ||   |       |   |  |   ||  |  ||  || || |||| ||
Db  73620 AATAATAATAAATATTAAACAATTAAATAAAATATACATAATTAATATTAATAAATAATT 73679

Qy    959 ttaatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaacta 1018
          | |||    ||      |||| |  |  |  |||||    || ||| ||| |   |  ||
Db  73680 ATTATATTAAAATAATTAAAAAAATTAATTAATTTAAGATAATATATAATATTATATATA 73739

Qy   1019 caaatccagatataatgg--------------tagatatgtacttatcccatttatattt 1064
            |   |   ||||  |              ||  ||   | |||    ||||   |||
Db  73740 ATTAATTATTAATAAATGTTTTATATTTAATTTAATTACACAATTAAATTATTTTATTTT 73799

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
            | | || ||    |    |   ||  |   ||||   |||    | |   ||  |
Db  73800 ATATATTAATTAAGTTAATTAATAAATGGTTATTTTTATTTTAATTAAAATGTGAAATAT 73859

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
          |||   |     |   ||| || |||    ||   ||   ||   | || ||| | | |
Db  73860 ATTATTTATTATATTAAAAAAAATATATATATATTTAATTAATTAATTATATGTTTATAT 73919

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
          |  || | | ||||      |   |||      |    | ||||||| ||| ||  |
Db  73920 TTTAATTTAAAATA------AATATATTTTATAAAATTAATTATTTATTAAATAAATAAA 73973

Qy   1245 accaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaatcc 1304
          |  || |     ||   ||    |||  |  | || || |||||   |  || ||  |
Db  73974 ATAAAATAATAAATATTATTATAATAAATAACATAAAATTATTTATTACTATATATTTAA 74033

Qy   1305 tgaagtacaaactccaaaagaccttactataactgaaaaccctgaagactgtagcacaat 1364
          | || || ||     |||  | ||| ||  |  |  |         | | |       ||
Db  74034 TTAAATAAAAGTATAAAATTATCTTTCTTAATTTATATTTTAAATTAATTATTAATATAT 74093

Qy   1365 agatgaatatgttccaggactcgttattccaaccatagggatatattacaaacacaatga 1424
          | ||  |  | |   |  | |  |||||  ||    |  |||   |   |  | | | |
Db  74094 ATATATAATTATAAAATAAATTATTATTTTAATTTAATTAATAATATTATATAATATTTA 74153

Qy   1425 tgctatatggaaaa 1438
          |  ||| ||   ||
Db  74154 TATTATTTGTTTAA 74167


RESULT   9
AC004157
LOCUS       AC004157   169546 bp    DNA             HTG       12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AC004157
VERSION     AC004157.8  GI:9797712
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 169546)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169546)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810447.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    23466: contig of 23466 bp in length
            *   23467    23666: gap of unknown length
            *   23667   169546: contig of 145880 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169546
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC293"
                     /clone="3D7"
BASE COUNT    69871 a  15381 c  15705 g  68389 t    200 others
ORIGIN


  Query Match             5.8%;  Score 102.8;  DB 2;  Length 169546;
  Best Local Similarity   45.5%;  Pred. No. 3.3e-06;
  Matches  662;  Conservative    0;  Mismatches  757;  Indels   35;  Gaps    7;

Qy      9 taaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaacctt 68
          ||| | ||  | |  | | | || | | |  || ||  ||     ||| |     |  ||
Db  84309 TAATAAATAGAAAAATATTAATACAATTTAAATATTAAATAAATAAAAATATTATAATTT 84368

Qy     69 gaaaggaaaaaataatataaataattcattgggaataatacgggaaaata--aaaataaa 126
            ||  || ||| ||  | |||| | | |    |  |||    ||| ||  | ||| ||
Db  84369 ATAAATAATAAAATATTAATATAAATTAATTAATAATATATAATAAATTAATATAATTAA 84428

Qy    127 actcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaatttt 186
          | | |  | |   |   | |  |   |  | |  || ||||    ||  | | |||| |
Db  84429 ATTTAAAATATAAATTAATAAAAAAATAATACTAATATTAATATAAAATAAAATAATAAT 84488

Qy    187 gtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaatataaaa 246
            | ||  |  |  | |       | |       ||    || || |  |  |||||||||
Db  84489 AAATAAATTTTAATTAAAATTAAATAATAAAATATTAATATAAATTAATTAAATAATAAT 84548

Qy    247 aatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaaagg 306
          |  |||||| ||     |||  ||    |  ||   || |  |  |  |  | ||| ||
Db  84549 ATAATAAATTAATTAAATAATAATATAATAAATTAATTAATTAACAATATTTAAATAATT 84608

Qy    307 aaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaa 366
          ||| |   || ||||      |  .   |  ||  |  ||      |  | ||  |  | ||
Db  84609 AAATATAATAATATATATTAAACAATTAATTATTATAAATTAAAGAATTATTAATAATAA 84668

Qy    367 catatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaat 426
          ||||  |  |  | |||   ||||   |   |      ||       |  |     |||
Db  84669 TATATAAATTAATTAAATAAAATGAATTATTTAAATAATTAAAACAATAATATATATAAA 84728

Qy    427 atgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacac 486
          |  |      || ||   |    | | ||   || |   | |||  || |    |   |
Db  84729 TTAATTATATATTTAGTAAATAAAATAAAATTAATAATTAAATTAATAATTTATAGTTAT 84788

Qy    487 ggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagta 546
              | || |  |||  |||  |   ||   |||  |||||| |   ||    |||  |
Db  84789 AAAAAATAAAAATAAATATATAATTAAATATATAAATAAATTAAATAAATATATAAATAA 84848

Qy    547 cgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatatagaa 606
            |  | || ||   |    || |      |  |  | | |  || || | | |||||  |
Db  84849 ATTAAATATATATACA---ATTAAATTAAATATATATATAATTAAATTAAATATATATAT 84905

Qy    607 aaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggt 666
          ||  ||||| | ||| |  ||||    | | ||  |  | ||| |   |  ||| | |
Db  84906 AATTAAATTAAATATATATATAATTAAATTAAATTAATAAATAAAATAAATTAATAAATA 84965

Qy    667 gtagat---attgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
            | ||   ||| | |   |  | ||   |      |  |  |  ||  || | || | |
Db  84966 AAATATATAATTAAATTAAATAAATATATAATTAAATTAATATATAAATAAATTAAATAT 85025

Qy    724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
            | || |||| |||| ||||  ||| ||  ||| | | | |||      |  |||   ||
Db  85026 ATATATAATTAAATTATATTATATAAATTAATAATATATAATATAATAAATAAAATAATA 85085

Qy    784 atttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttc 843
          |||  || |  |  |  || | ||  |  || | ||    ||  |  | |   || ||
```

```
Db  85086 ATT--AAATAATAATATATAATATAATAAATAAACAATAATTAAATTAATAATATATTAT 85143

Qy    844 tgtaaagatgaagaatctccatataacactaaatttttgtctgaacaa-----atagaaa 898
Db  85144 AAATAATATAATAATTAAATATAATATAATAATTAATTAAATTAATAATATATTAAATAA 85203

Qy    899 caaataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattt 958
Db  85204 AATAATAATAAATATTAAACAATTAAATAAAATATACATAATTAATATTAATAAATAATT 85263

Qy    959 ttaatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaacta 1018
Db  85264 ATTATATTAAAATAATTAAAAAAATTAATTAATTTAAGATAATATATAATATTATATATA 85323

Qy   1019 caaatccagatataatgg--------------tagatatgtacttatcccatttatattt 1064
Db  85324 ATTAATTATTAATAAATGTTTTATATTTAATTTAATTACACAATTAAATTATTTTATTTT 85383

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
Db  85384 ATATATTAATTAAGTTAATTAATAAATGGTTATTTTTATTTTAATTAAAATGTGAAATAT 85443

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
Db  85444 ATTATTTATTATATTAAAAAAAATATATATATTTAATTAATTAATTATATGTTTATAT 85503

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
Db  85504 TTTAATTTAAAATA------AATATATTTTATAAAATTAATTATTTATTAAATAAATAAA 85557

Qy   1245 accaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaatcc 1304
Db  85558 ATAAAATAATAAATATTATTATAATAAATAACATAAAATTATTTATTACTATATATTTAA 85617

Qy   1305 tgaagtacaaactccaaaagaccttactataactgaaaaccctgaagactgtagcacaat 1364
Db  85618 TTAAATAAAAGTATAAAATTATCTTTCTTAATTTATATTTTAAATTAATTATTAATATAT 85677

Qy   1365 agatgaatatgttccaggactcgttattccaaccatagggatatattacaaacacaatga 1424
Db  85678 ATATATAATTATAAAATAAATTATTATTTTAATTTAATTAATAATATTATATAATATTTA 85737

Qy   1425 tgctatatggaaaa 1438
Db  85738 TATTATTTGTTTAA 85751


RESULT  10
PFMAL1P3
LOCUS       PFMAL1P3    67970 bp    DNA             INV       15-DEC-1999
DEFINITION  Plasmodium falciparum MAL1P3, complete sequence.
ACCESSION   AL031746
VERSION     AL031746.9  GI:6594243
KEYWORDS    HTG.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 67970)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Dec 16, 1999 this sequence version replaced gi:5763807.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.

FEATURES             Location/Qualifiers
     source          1. .67970
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="1"
     gene            complement(join(1748. .2598,2748. .2848,2990. .3276))
                     /gene="MAL1P3.01"
     CDS             complement(join(1748. .2598,2748. .2848,2990. .3276))
                     /gene="MAL1P3.01"
                     /note="MAL1P3.01, conserved hypothetical protein, len: 412
                     aa, similarity: UPF0006 family eg to
                     YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
                     scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
                     overlap)"
                     /codon_start=1
                     /product="conserved hypothetical protein, UPF0006 family"
                     /protein_id="CAB63556.1"
                     /db_xref="GI:6594244"
                     /translation="MKLVFHYIKYINVLFYISIIFLKSNSLKIYNDLRYISTVNKYKV
                     LQIKKRSNLKKNHNIRKMEDNESSFIDIGSNLTDKMFDGVYNSKKHENDLQNVLNRAK
                     NNNVDKIIITCTCLAEIDKSLKICETYDPEGKFLYLSAGVHPTNCYEFIDKNKHEEKE
                     IIAKKEYEEFIKYFKNEQVENSKMENGNKKICDGEKDMNNLNEILLEKNLDTIPGFKY
                     NEKDKEYLENLKNKIIKYPNRIVCIGEIGLDFDRLYFCSKYIQIKYFIFQLKLVQMFN
                     LPMFLHMRNCSETFFKIVDIYKFLFEKNGGVIHSFTDKEDIVHIIVQNYKNLYIGVNG
                     CSLKSLENINAVKKIPLNLLLLETDAPWCGVKKTHASYEYIKDTYEKRAYTNLKKIKN
                     IIKCDDNTIFKERNEPYNIA"
     misc_feature    complement(2599. .2610)
                     /note="potential splice acceptor sequence"
     misc_feature    complement(2742. .2747)
                     /note="potential splice donor sequence, atg/gttaaa"
     misc_feature    complement(2849. .2861)
                     /note="potential splice acceptor sequence"
     misc_feature    complement(2984. .2989)
                     /note="potential splice donor sequence, aaa/gtaaaa"
     gene            5005. .5496
                     /gene="MAL1P3.02"
     CDS             5005. .5496
                     /gene="MAL1P3.02"
                     /note="MAL1P3.02, hypothetical protein, len: 163 aa,
                     contains possible signal sequence"
                     /codon_start=1
                     /product="hypothetical protein, MAL1P3.02"
                     /protein_id="CAB63557.1"
                     /db_xref="GI:6594245"
                     /translation="MKLLNNRFVVLCPIIILFFFLNSVVLGNNNRNNINFHETENAAK
                     AMRKLLSGEINSIKLDNGDELKIKLNDEKHKDSTKWDKSYSFISNLEEEKYSQTDLFR
                     KKQEINEANTKIIEDRQEFYILNNDEIENIATRFVLENNFDELYIQSFKQSLIDIIQS
                     LNN"
     misc_feature    8020. .10389
                     /note="possible cen1, region of very high [A+T] content"
     gene            14884. .20352
                     /gene="MAL1P3.03"
     CDS             14884. .20352
                     /gene="MAL1P3.03"
                     /note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
                     /codon_start=1
                     /product="putative ABC transporter"
                     /protein_id="CAB63558.1"
                     /db_xref="GI:6594246"
                     /translation="MTTYKENVGISNKGNKKKKSCQNISFLNFLSFDWIRPLINDLIK
                     GDIQELPNICRNFDVPYYASKLEENLRDIEVEDSEFYSEKNSSNEHVLHHCNSNDASE
                     KKVYNVYYHNILWSILKTFKFRIILIISFYILETLIVTLGGKFIDYYMRILEGQKIPV
                     YISFLKDFKVFSGLVVVMIMFFHLFFEALLHFYFHLFTINLKVSLMYFLYKINLCSNN
                     NHLQNPDAFYNTYRKFSSQTEIDEISRDFLSIGKNASSSSSGIKNNNKNIDNNKFVEN
                     DYIINFIKSTKKMEKDSLNENRSLPNVNIYNIMFSDVPSVTFFVTSCINLFNVFVKIF
                     MSFYVFHIKIGSNSVGIAIWLSIALYSAMILFEFLPSLFKSKYLIYRDKRIDNMHHVL
                     KEFKLIKMFNWESFAFKYINIFRMKEMKYCKIRLYLSNIGVFISSISSDIVEVVIFFI
                     YLKDRLNKKEEIKFTSIIMPLYVYKILISNVANFPNLVNNVMEGIVNIKRLNNYINDH
                     LYYNDIKNYFMYRTRYNEDYNIVVDKTFLQNENITSHDDGTSHNLKHLKNVIKNKLTN
                     MFKYFFFYHKMNYHKNIINKQILSGLLKNVDDNTNKKICFQEHKSNSTYNYNSSHIHE
                     KKEEYENIHNSSNSTMSNEFKEKKKNNEYIIKLENCSFGLSYDNKCDNDHILKNINFN
                     LKRNSLAIIIGNVGSGKSAFFHSILGDFNMTHGNLYIENFFKKMPILYVPQNSWLFMG
```

```
                     NIRSMILFGNEYNPLIYKYTILQSELLNDLSTIEHGDMKYINDDHNLSKGQKVRICLA
                     RALYEHYIHMHKLCTDYEKKLIQPNEILDKDLINNKNISSYNNKKSKLVNYNIPFNEN
                     YLQKCLMDDNNFYLYLLDDIFTSLDPSISKKIFSNLFCKEDNISFKDNCSFIISMNKS
                     TLDNFLIEDILDNVQYEVNIFEIQDKTLKYRGNISEYMEKNNLNITKESHWGYSNLNT
                     IDYTRIKLFDEVELNHVKHSNKMIYKEAYFVKGNTESVSFEIDSINKEYIKKMKKKNY
                     KKEHMNKNNKDNNNNNNNSNKDDHININMNDNHRNYNDINLGPNSTDDSPTVSSLGNE
                     YTLDTYTSNNSDKEEIVKPLYKDTHEEFNKSSSMPFVKSSSNMINNPSNFKYEDNSSS
                     FKGSISLETYLWYFQQVGFVLLTSVVIFMLISIFTDEIKFVFLTMMSIISKNNKEHSD
                     TILQKQVRYLEYFVILPIISLVTSGICFSMIIYGNITSAIKVHNNILYSILNAPLYIF
                     YNNNLGNIINRFIIDISAFDYGFLKRIYKAFFIFFRCILSSLLIIYMIRDCIFIFPFV
                     IILIYFFVFKRFSRGCKEAQRLYLSCHTPLCNIYSNALSGKNIINIYKKNTYHLDVYE
                     HYINNFRISYFFKWLINIWASLYIKIFILLLTTYIIMHPHLYASGIIKLYKEKNYVRI
                     LSTLGYCISFSARLGVIIKFLLCDYTHIEKEMCCVQRLEEFAKISNKENASMNKENEL
                     NVITTQTYKEKNENISDKISAIVEYKNVSLSSIINSSQDDESKKKYGIKFENVYVSYK
                     KKIPLVNGTYKYIDEEPSLKNINMYALKNQKIGIVGKSGAGKSTILLSILGLINISQG
                     KITVEGRDIRTYNRKGEDSIIGILAQSSFVFYNWNIRTFIDPYNNFTDDEIVHALKLN
                     GINLGKNDLYKYMHKQDMKSNYKKIIQTSKVINQSNDNTILLTNDCIRYLSLVRLYLN
                     RHKYKIILIDEIPIFNLNNSVHDELNSFLIGKAKSFNYIIRNHFPNNTVLIISHHANT
                     LSCCDYIYVLRKGEITYRCSYEDVKTQSELSHLLEMDD"
rRNA                 23896. .31533
                     /gene="rRNA"
                     /note="region containing small subunit, 5.8S and large
                     subunit rRNA genes and spacer regions"
gene                 23896. .31533
                     /gene="rRNA"
gene                 complement(join(31966. .32476,32675. .32775))
                     /gene="MAL1P3.04"
CDS                  complement(join(31966. .32476,32675. .32775))
                     /gene="MAL1P3.04"
                     /note="MAL1P3.04, conserved hypothetical membrane protein,
                     len: 203 aa, similarity: P. falciparum chromosome 2,
                     PFB0110W, O96126 predicted integral membrane protein (255
                     aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
                     in 191 aa overlap)"
                     /codon_start=1
                     /product="conserved hypothetical membrane protein,
                     MAL1P3.04"
                     /protein_id="CAB63559.1"
                     /db_xref="GI:6594247"
                     /translation="MKKSYTFINVTILLFLTLLLFLTYYNYDTFSKTKFNNNIKIDIN
                     RFKRIIAEASEEQKYPWEEDFCLILNEEELIRPEHNDSPYLPEHYENIDKINELSINS
                     TKIWKETIKKMRQNYEKETDNMNHNWRDFMWHYKWANIYLYKVHKLINITLKDLTNPI
                     HDKEETITTWIKWIQEDIEYFLFNLQVEWLRILTLELFYKNKE"
misc_feature         complement(32477. .32486)
                     /note="potential splice acceptor sequence"
misc_feature         complement(32669. .32674)
                     /note="potential splice donor sequence, aaa/gtatat"
gene                 join(36657. .36743,36864. .37343)
                     /gene="MAL1P3.05"
CDS                  join(36657. .36743,36864. .37343)
                     /gene="MAL1P3.05"
                     /note="MAL1P3.05, hypothetical protein, len: 188 aa"
                     /codon_start=1
                     /product="hypothetical protein, MAL1P3.05"
                     /protein_id="CAB63560.1"
                     /db_xref="GI:6594248"
                     /translation="MRIKMNSGIFFIKLLICISFICVFECFNKCMISYRKDLLWYSEN
                     CFNYSIDRSLAEGSSESKETKVKDIPNIELLKSLNINYEEYEKMKEIVGSFMDNNNLN
                     IANEVLKNIHSFTNIENIFSLINDSSKSPVLKTFLKEFGSIFPHMLNNVPKLLFDLCQ
                     RNPLHIILGLIVILAAIYVFENFKNFEC"
misc_feature         36744. .36749
                     /note="potential splice donor sequence, aag/gtatga"
misc_feature         36854. .36863
                     /note="potential splice acceptor sequence"
gene                 complement(join(38049. .39995,40210. .40284))
                     /gene="garp"
CDS                  complement(join(38049. .39995,40210. .40284))
                     /gene="garp"
                     /note="MAL1P3.06, garp, len: 673 aa, similarity: almost
                     identical to GARP_PLAFF (678 aa), fasta scores: 97.6%
                     identity in 678 aa overlap"
                     /codon_start=1
                     /product="hypothetical garp protein"
                     /protein_id="CAB63561.1"
                     /db_xref="GI:6594249"
                     /translation="MNVLFLSYNICILFFVVCTLNFSTKCFSNGLLKNQNILNKSFDS
                     ITGRLLNETELEKNKDDNSKSETLLKEEKDEKDDVPTTSNDNLKNAHNNNEISSSTDP
                     TNIINVNDKDNENSVDKKKDKKEKKHKKDKKEKKEKKDKKEKKDKKEKKHKKEKKHKK
                     DKKKEENSEVMSLYKTGQHKPKNATEHGEENLYEEMVSEINNNAQGGLLLSSPYQYRE
                     QGGCGIISSVHETSNDTKDNDKENISEDKKEDHQQEEMLKTLDKKERKQKEKEMKEQE
                     KIEKKKKKQEEKEKKKQEKERKKQEKKERKQKEKEMKKQKKIEKERKKKEEKEKKKKK
                     HDKENEETMQQPDQTSEETNNEIMVPLPSPLTDVTTPEEHKEGEHKEEEHKEGEHKEG
                     EHKEEEHKEEEHKKEEHKSKEHKSKGKKDKGKKDKGKHKKAKKEKVKKHVVKNVIEDE
                     DKDGVEIINLEDKEACEEQHITVESRPLSQPQCKLIDEPEQLTLMDKSKVEEKNLSIQ
                     EQLIGTIGRVNVVPRRDNHKKKMAKIEEAELQKQKHVDKEEDKKEESKEVEEESKEVQ
                     EDEEEVEEDEEEEEEEEEEEEEEEEEEEEEEEDEVEEDEDDAEEDEDDAEEDEDDAEED
                     DDDAEEDDDDAEEDDDEDEDEDEEEEEDEEEEEESEKKIKRNLRKNAKI"
misc_feature         complement(39996. .40005)
                     /note="potential splice acceptor sequence"
misc_feature         complement(40204. .40209)
                     /note="potential splice donor sequence, aag/gtaaca"
gene                 join(45401. .46396,46562. .50233)
                     /gene="MAL1P3.07"
CDS                  join(45401. .46396,46562. .50233)
                     /gene="MAL1P3.07"
                     /note="MAL1P3.07, hypothetical protein, len: 1555 aa"
                     /codon_start=1
                     /product="hypothetical protein, MAL1P3.07"
                     /protein_id="CAB63562.1"
                     /db_xref="GI:6594250"
                     /translation="MEQIGIANNIFTEIGKNIVKRSGPFEIWRKKFIEEVSKKISNSL
                     NIENIKECKKNCRKINYWMDDKEEEFVSKKFFVYEYITSNVWKNEFEKAIINTLKNKT
                     KTCLRSRKKYPKEIRNIMGELEDYIDAINDYKKQFKNLYCWSERYIDYKKWLNEMKEY

 Query Match             5.8%;  Score 102.6;  DB 3;  Length 67970;
 Best Local Similarity   44.9%;  Pred. No. 4e-06;
 Matches  650;  Conservative    0;  Mismatches  779;  Indels   20;  Gaps    6;

Qy       9 taaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaacctt 68
           ||||||||| || || | ||| |  |  ||  |  |  ||      | |||  |  |  |
Db    8240 TAAAATATTAAAATATATAAATTAATAATATAATTAATAAATAAAATATTATATTAAATA 8299

Qy      69 gaaaggaaaaaatatatataaataattcattgggaataatacgggaaaataaaaataaaac 128
             |    | ||||| || ||   ||  ||| |||| |    ||||||||    |
Db    8300 AATTAAATAAAATATTAAAATATATAAATTAATAATATAATTAATAAATAAAATATTATA 8359

Qy     129 tcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaatttttgt 188
           |     ||  || || || |    ||   | || | || |   |  ||| | | |   |
Db    8360 TATATTAAATAAATTAAATAAAATATTAAAATATATAAATTAATAATATAATTAATAAAT 8419

Qy     189 agaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaaaa 248
           | |||||      ||||   | ||      || | |  || || | | ||| ||||| |
Db    8420 AAAATATTA----TATTAAATAAATTAAATAAAATATTAAAATATATAAATTAATAATAT 8475

Qy     249 tataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaaggaa 308
           |||| |||  ||| |  || |     |  || || | |  |  || || |||  ||
Db    8476 ATATAATTAAATAAATAAACTATTATATTAAATAAATTAATAATATAATTAATAAATAAA 8535

Qy     309 acaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaaca 368
           | |  |||||         | ||   |  ||  || | | ||    ||| |  || |
Db    8536 ATATTATATTA---------AATAAATTAAATAAAATATTAAAATATATAAATTAATAAT 8586

Qy     369 tatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatat 428
             ||   |  | |   ||      |  |||  ||| |||    |  |   |||||||
Db    8587 ATATATAATTAAATAAATAAAATATTATATTAAATAAATT---AATAATATAATTAATAA 8643

Qy     429 gttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacgg 488
           | | ||  | |||  |    ||| |||  ||  |  || ||   ||     | || |
Db    8644 ATAAAATATTATATTAAATAAATTAAATAAAATATTAAAATATATAAATTAATAATATAA 8703

Qy     489 cttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacg 548
           | | ||    ||  |||          | ||| |||| | | ||   |  | | |
Db    8704 TTAATAAATAAAATATTATATATATTAAATAAATTAAATAAAATGTTAAAATATATAAAT 8763

Qy     549 tccagatgtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaa 608
           |    || ||||||   | ||||  | |       |||   |         || | ||
```

```
Db   8764 TAATAATATAATTAAAATATTATATTAAATGAATTAAATGAAATATTATATTAAATAATT 8823

Qy    609 agaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggt-g 667

Db   8824 AAAATATTATTATATTAAATAAATTAAATATTATATTAAATAAATAAAATAATATATTAA 8883

Qy    668 tagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaa 727

Db   8884 AATATATAAATTAATAAATAAAATAATAATATAATTAAATAATTAAATAAAATAATAATA 8943

Qy    728 attattatattaaattaattaacttgttaagaaaaactattccggaagaaaagttaattt 787

Db   8944 TATTAAATAAATAATTAAATAAAATAATA--ATATATTAAACAAATTAATAAATAAAATA 9001

Qy    788 caatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttctgta 847

Db   9002 AAATATTATATTAATTTATATGTTTAATATATTGTCTTTATTATTTAATATAATATTTTA 9061

Qy    848 aagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaaataaag 907

Db   9062 AATAAATAAAAATATATTAATATAATAATTTAATAAATAAAACAATTTATATTATATTAA 9121

Qy    908 aattacatagggcagcagcgatgttatcagcaggaacttttattaatatttttaatacag 967

Db   9122 ATAATTAAATATATAAAATTAATATAATAATTTAATAAATTAATTAAATTATTTAATATA 9181

Qy    968 caaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactacaaatccag 1027

Db   9182 TAAAATAAATATATACGCTAATAAATAATTAAATATATAAAAATATAATAATAAATAAAA 9241

Qy   1028 atataatggtagatatgtacttatcccatttatattttggtttaaaatataacatcacaa 1087

Db   9242 TTAAAATTATTAAATTAATTAAATAATATAATAAATTAATTATTTAATATAATTAATATA 9301

Qy   1088 tcat-attaggtttttcattagaacataacagaggtggatttagtcccgaaaataaagaa 1146

Db   9302 TAATAATTAATTTATTTATTTTAGTAACATTATAATAAAATAATATATATATAATAAA 9361

Qy   1147 ttattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagggcagat 1206

Db   9362 TTAATATATTTATTAAAATAAAAACAAATATATATTTATTATATTTTTAATATATAAATA 9421

Qy   1207 ggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattcgatgta 1266

Db   9422 TTTTCATTTTATTTATTATATTTTTATTAATTATATTTTTAAATAATAAATTACACATTT 9481

Qy   1267 gatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactccaaaagac 1326

Db   9482 AATTAATTATTATTTATTTAATTATAACTTTATAATATTTAAATATAATATTAATTATTT 9541

Qy   1327 cttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttccaggactc 1386

Db   9542 TATATTTTTTAAATATATATATAAATAATTATAATAATATATTTTAATTAATTAAATAAT 9601

Qy   1387 gttattccaaccatagggatatattacaaacacaatgatgctatatggaaaactagatct 1446

Db   9602 TTATATTTAAAATGTGATATATATATATATATATATATTTATATATGTAATATAATTAAA 9661

Qy   1447 tattcaatt 1455

Db   9662 TATTAAATT 9670


RESULT  11
AC090493/c
LOCUS       AC090493   213530 bp    DNA             HTG       25-JUL-2001
DEFINITION  Mus musculus clone RP23-201D19, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION   AC090493
VERSION     AC090493.2  GI:15011687
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 213530)
  AUTHORS   McCombie,W.R., Baker,J.P., Bahret,A., Yang,C., Balija,V.,
            Dedhia,N.N., de la Bastide,M., Kuit,K., King,L., Kirchoff,K.A.,
            Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,
            Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Palmer,L.,
            Vil,M.D. and Zutavern,T.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213530)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2001) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     On Jul 25, 2001 this sequence version replaced gi:13122744.
            ------------------- Genome Center
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact:  mccombie@cshl.org
            ------------------- Project Information
            Center project name: RP23-201D19
            Center clone name: RP23-201D19
            ------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   100656: contig of 100656 bp in length
            *  100657   102023: gap of unknown length
            *  102024   172443: contig of 70420 bp in length
            *  172444   173810: gap of unknown length
            *  173811   210022: contig of 36212 bp in length
            *  210023   211388: gap of unknown length
            *  211389   213530: contig of 2142 bp in length.
FEATURES             Location/Qualifiers
     source          1. .213530
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-201D19"
BASE COUNT    54181 a  51263 c  51890 g  51971 t   4225 others
ORIGIN


  Query Match             5.5%;  Score 97.6;  DB 2;  Length 213530;
  Best Local Similarity   43.1%;  Pred. No. 1.7e-05;
  Matches  519;  Conservative    0;  Mismatches  666;  Indels   20;  Gaps    6;

Qy     11 aaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaaccttga 70

Db 209584 AAATATATATTATAACATNTATTNNATACTTAAATATAANTTAAAAAAAAAAAAAATAAA 209525

Qy     71 aaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaataaaactc 130

Db 209524 ATAAAAATATTATANAANATAAAATATTTTATATAAAAATTTTAAATAANATTTNATATA 209465

Qy    131 atcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgtag 190

Db 209464 TATTAAAAAATAATTTNNAAATTNTTAAATATAAAAANNAAAAAAATATATTTAATATAA 209405

Qy    191 aatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaaaata 250

Db 209404 AAATTATATATAATNACNNAAANTTTATTAATT------TTTAAAANTNANATAAAAATT 209351

Qy    251 taaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaaaggaaac 310
```

```
Db 209350 TAATTAAACAAAAAAAAAAATAAANNAAANAATTTAATATAAATAATCAAAACAATATAT 209291

Qy    311 aaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacata 370
          |    | ||||    |||       |   | |          | ||| |  |||| | |
Db 209290 ATATAAATATACTTTTATAAATNCATNTNTAAATATTAATTTTTNATAAAAAAAAAAAAA 209231

Qy    371 tgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgca---ttaata 427
          |||  ||    |   ||        |    | ||||   || |   | ||    ||  |
Db 209230 ATATTTATAATCATAAAAAAANAAACAAACAAAATATAAAATCTAAACCCAAACCTATAA 209171

Qy    428 tgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacg 487
          | |||| | |   |   | |  | || ||  |  | |||  |    || || |||| |
Db 209170 TATTATTTTAACAACTAAAATAAATTTATTNATAAAAAAATTATNAATAAAAAAAAAATT 209111

Qy    488 gcttagaatatgaaacctatggtatgatgctta----atgaaattagacgtatcagaaaa 543
          |  ||||   | |    ||  || |  |    ||  ||  |    | | | ||||
Db 209110 TAAAAATATATANCAACNNATATAAAATACAAAAAATATTTAANAAATTTTTTTAAAAAA 209051

Qy    544 gtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatata 603
          |  | || |    |       ||   || |   |||   | ||     || |  | |
Db 209050 ACTANTATATATATTTAAAANNAAAAATAAATAAATTATAAAANANAAATATATATTAAA 208991

Qy    604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagat 663
          || |  ||   ||| |    | |||| |       ||   ||        || |  |
Db 208990 TTAATANNAAAANATTTTTATAAAAAAAAACAATTTATTAAAAAATNAAANCATATCAAN 208931

Qy    664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
          | |||||  |     ||  |      || ||    |   |||| ||  | |   |
Db 208930 AANTAAAATATTTAAANAAAAAAAATAAAAAAATTNNNTATAATAAAAAAAAAANANANT 208871

Qy    724 tcaaattattatattaaattaattaacttgttaagaaaa--actattccggaagaaaagt 781
          ||  || |||||  |||  | | || |    || ||||  |  |||        |||
Db 208870 CAAANATAATATATAAAAAAATTAAAATAAAAAAAAAAAAATATAATTTATTTTTCAAAAA 208811

Qy    782 taatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctt 841
          |||  | ||       | || ||||     |||  |  |         |  |      |  |
Db 208810 AAATNTAAANNNNCNAATAATAAAAT----AATTNTAAAATAAAATAAAAAAAAAAAATN 208755

Qy    842 tctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaa 901
          | |||| |  |  |||       |||  |  |||   |  |    | |   | ||  ||
Db 208754 AATATAAAAAATATAAATANTATAATANNAAAAAAAANTATTAATTATANTAAAAANAAA 208695

Qy    902 ataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttta 961
          || |  || |  |       |  |   |  |   ||  |    |       |||| |  | |
Db 208694 ATTATAAANTCTANNCANNAAAATAAAAATATACATTAAATAANAAATATAATNTAATAA 208635

Qy    962 atacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactacaa 1021
          |||  |  | |  | |||| | |  || | |  ||  |  |||| || | |  || || ||
Db 208634 ATAAAAAATATTACAAAA-ACANATTTTTTANATAAATCCATATAAATAAATAAATANAA 208576

Qy   1022 atccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataaca 1081
          ||  |  ||||     |  |||  |  ||| ||| |  | |       |  || | | || |
Db 208575 ATTAAATTATATATAAAATTATAAAAATATACCAAAAAAAAAAAAANTNNAATAAAAAATA 208516

Qy   1082 tcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaata 1141
          |   | | | | ||              | |    | || |     ||  |      ||||
Db 208515 TATNANTAAAAATAATAAANNNTNNAAATTNAACCAAAAAAATATAAAANATTTAAAAAN 208456

Qy   1142 aagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataataggg 1201
          |||| |       ||  | ||  |  |      ||||  |||| ||| | |
Db 208455 NCNTATTAAAATCCAAAATANAANAANTATAAACAACAAAAAANTAAAAAAATTAAATTA 208396

Qy   1202 cagat 1206
          || ||
Db 208395 CATAT 208391


RESULT  12
AC012492/c
LOCUS       AC012492   183584 bp    DNA             PRI       20-MAY-2001
DEFINITION  Homo sapiens clone RP11-224B10, complete sequence.
ACCESSION   AC012492
VERSION     AC012492.9  GI:14151030
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183584)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183584)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 183584)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 20, 2001 this sequence version replaced gi:13431121.
            Center project name: H_NH0224B10.
FEATURES             Location/Qualifiers
     source          1. .183584
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-224B10"
BASE COUNT    53876 a  33878 c  34014 g  61816 t
ORIGIN


  Query Match             5.3%;  Score 93.8;  DB 9;  Length 183584;
  Best Local Similarity   44.4%;  Pred. No. 6.2e-05;
  Matches  558;  Conservative    0;  Mismatches  687;  Indels   12;  Gaps    4;

Qy      7 tttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaacc 66
          || || |||   |||||| ||||      ||   | |   | | |  ||  |  |  |
Db  16436 TTAAATATAAATATATTTATAATATATAAATATTTATATAAATATAAAATATATATTAAA 16377

Qy     67 ttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaataaa 126
          |  |||   | | | |||||| || |   ||    ||||  |    | | |||| ||| |
Db  16376 TATAAATATATATAAAATATATTAAATATATAAAATATAAATATATATTAAATATATA 16317

Qy    127 actcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaatttt 186
              ||  |    | ||| ||   | ||    | || ||||||  |  ||   ||| | |
Db  16316 TTAAATATATAAAATATAAATATATATTAAATATATTTTAAATATATAAAATATAAATAT 16257

Qy    187 gtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaatataataa 246
          ||  | ||  || |||      |   |  | | ||  |     |     |||| |
Db  16256 ATATTAAATATATTTAAATATATTAAATATAAATATATATTAAATATATTTTAAATATA 16197

Qy    247 aatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaagg 306
            |  |  || |   |||  ||||    |  | | || | | |    || || | ||
Db  16196 TTAAATATAAATACATATATTAAATATATATTATATATATAAAATATATAAAATATAAAT 16137

Qy    307 aaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaa 366
          | | |   ||      |  |||     ||| |    | |    |   |  |  |||
Db  16136 ATATATTAAATATATATAAAATATATATGTTAAATATATAAAAGATATATAAAATATAAA 16077

Qy    367 catatgattgattcaaacccaatgg-----tgtcaattttatatattgcatttgctcgca 421
          ||||  |  ||  | |   |||       | | || | ||||||||  || |      |
Db  16076 TATATATTAAATATATATAAAATATATATATTAAATATATATATTAAATATAAATATA 16017

Qy    422 ttaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaa 481
          |  | |   || ||        || |    |  ||  || | ||| ||   ||||||  |
Db  16016 TATAAAATATAAATATATGTATTAAATATATATATTAAATATAAATATATGTATTAAATA 15957

Qy    482 aacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaa 541
          | |    |  ||||||||        |   ||   || | ||  || | |  |||
Db  15956 TATA-----TTAAATATGAATATATGTATTAAATATATATTAAATATAAATATATGTATT 15902

Qy    542 aagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagata 601
```

```
Db  15901 ATATATATAGAATATAAATATATGTATTAAATATAGTATATTAAATATAAATATATATAA 15842

Qy    602 tagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttag 661
Db  15841 AATATATATTAAATATGAATATATATAAAATATATATATTAAAAATATATATAATATAAA 15782

Qy    662 atggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatt 721
Db  15781 TATATATAAAATATATATATTAAAAATATATATAATATAAATATATATAAAATATATATA 15722

Qy    722 tttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagt 781
Db  15721 TTAAAAATATATATAAAATATATATATTAAAAATATATATAAAATATATATATTAAAAAT 15662

Qy    782 taatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctt 841
Db  15661 ATATAT-AAAATATATATATTAAAAATATATATTAAATATAAATATATATATTAAAAATA 15603

Qy    842 tctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaa 901
Db  15602 TATATTAAATATAACTATATATTAAATATATATTAAATATAACTATATATTAAATATATA 15543

Qy    902 ataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatatttta 961
Db  15542 TTAAATATAACTATATATTAAATATATATTAAATATAACTATATATTAAATATATATTAA 15483

Qy    962 atacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactacaa 1021
Db  15482 ATATAACTATATATTAAATATATATTAAATATAACTATATATTAAATATATATTAAATAT 15423

Qy   1022 atccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataaca 1081
Db  15422 AACTATATATTAAATATATATTAAATATAACTATATATTAAATATATATGAAATATAACT 15363

Qy   1082 tcacaatcatattaggtttttcattagaacataacagaggtggattta-gtcccgaaaat 1140
Db  15362 ATATATTAAATATATATTAAATATAACTATATGTATTAAATATAAATATATGTCTTAAAT 15303

Qy   1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200
Db  15302 ATATATTAAATATAAATATATGTATTAAATATATATTAAATATAAATATGTGTATTAAAT 15243

Qy   1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatca 1257
Db  15242 ATATATTAAATATAAATATGTGTATTAAATATATATTAAATATAAATATGTGTATTA 15186


RESULT 13
PFMAL3P5/c
LOCUS       PFMAL3P5    86827 bp    DNA             INV       04-MAY-2000
DEFINITION  Plasmodium falciparum MAL3P5, complete sequence.
ACCESSION   AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
            AL010206 AL010210 AL139179
VERSION     AL034556.3  GI:7711064
KEYWORDS    HTG; centromere; CTRP protein; initiation factor E4;
            Serine/threonine protein phosphatase.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 86827)
  AUTHORS   Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
            Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T.,
            Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
            Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
            Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
            Rajandream,M.-A., Rutter,R., Skelton,J., Squares,R., Squares,S.,
            Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and
            Barrell,B.G.
  TITLE     The complete nucleotide sequence of chromosome 3 of Plasmodium
            falciparum
  JOURNAL   Nature 400 (6744), 532-538 (1999)
  MEDLINE   99376085
REFERENCE   2  (bases 1 to 86827)
  AUTHORS   Bowman,S., Skelton,J., Churcher,C., Lawson,D., Quail,M. and
            Barrell,B.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 86827)
  AUTHORS   Lawson,D., Bowman,S. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On May 14, 2001 this sequence version replaced gi:2982535
            gi:2982536 gi:2894454 gi:2982554 gi:2982562 gi:2894489 gi:2982572
            gi:2982574 gi:4493931.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .86827
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="MAL3P5"
     gene            join(322. .603,826. .1023,1197. .1301,1458. .2942)
                     /gene="PFC0575w, MAL3P5.1"
     CDS             join(322. .603,826. .1023,1197. .1301,1458. .2942)
                     /gene="PFC0575w, MAL3P5.1"
                     /note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689
                     aa, possible signal sequence, revised: added new exon 2"
                     /codon_start=1
                     /product="Hypothetical protein, PFC0575w"
                     /protein_id="CAB38969.2"
                     /db_xref="GI:7711065"
                     /db_xref="SPTREMBL:O97258"
                     /translation="MYLKNVYIYISSCFILFDLCFSFHLLKMKYKNHMNNMKSVTFFL
                     RSPQIYRKRFKRSRIKNVSFKKKQKKPLFLFENLKKGFSFLGFWRNQYDQKYIDDVIS
                     NINNLTRIKQVTHKKKSNEFTKENIKQILLHCVFSKIDFKIINNLSYIIKHFQMSNIT
                     VHSILNQISEKVKEKKDAENYLALHLFLLKDENITLFSMMHIMDFFKSKQKVIECIRD
                     IKSKKKRKKNLSIYINLFICTLIYFTYCMCLLIKYISHLCIFFFFFFCFFLCYNILER
                     IYEECVGDLIRKKIERYNLYCEKKKIKFHMKDAIKKMEINMKDDDLYFNYHYDELLRC
                     FTMKLNIERNNKNIIRSNYDNINNDISIDKDMYMNNPIDVNINNISLDEKIKEQFENP
                     DDENLKELKDTYEQFQLFNDNIIKYIEEDQPLYNINDNSNINDNNNNINTMKNKHKIK
                     DTYNDDDDYDYEKEEDLVIQKNIDDYIYKNTIGMNKSLEEFKNQFIEQADIEFQNFLS
                     NVNLDQHGRVKSNDENTKSTEHIKNKNTINKGYDTELIQNQMENNFIKKNIDNNISND
                     ISNNEINIIKLKKLNQSDEDINLTSDLIYERLRTKVLWYIQKIEYLKFKYQYDIINEQ
                     YPIIKNEKTVLDLLNYGYKIVMSPDVDNSLFEKTKIDSIPNEKDKNNQMENQKNSKNY
                     IIMFNVM"
     misc_feature    604. .609
                     /gene="PFC0575w, MAL3P5.1"
                     /note="predicted splice donor sequence for exon 1 of
                     PFC0575w (revised)"
     misc_feature    816. .825
                     /gene="PFC0575w, MAL3P5.1"
                     /note="predicted splice acceptor sequence for exon 2
                     (revised of PFC0575w)"
     misc_feature    1024. .1029
                     /gene="PFC0575w, MAL3P5.1"
                     /note="predicted splice donor sequence for exon 2 of
                     PFC0575w (revised)"
     gene            complement(3354. .6644)
                     /gene="PFC0580c, MAL3P5.2"
     CDS             complement(3354. .6644)
                     /gene="PFC0580c, MAL3P5.2"
                     /note="PFC0580c (MAL3P5.2), Hypothetical protein, len:
                     1097 aa, possible signal sequence, predicted using
                     hexExon"
                     /codon_start=1
                     /product="Hypothetical protein, PFC0580c"
                     /protein_id="CAB38968.1"
                     /db_xref="GI:4493932"
                     /db_xref="SPTREMBL:O97257"
                     /translation="MLGLKRKNVFYLLVSVPSLFAYFLKRHKDNENNYETLINNNDIE
                     KIKKIRIHNKCSYIPLLFLNIYDSYIYKNKILRWLYFKFRKRRKDKEEYYYITNMVRK
                     KRREAIKYNFISDEQNLFNKFYIYEIVLEYSLKYGILSPHLSLYILKNISEHCVNIYP
```

```
                SLYYYNKLDNKHNLINEKKLKYFKQINNEHTQQAPTNHTHHNNNNNNKKPLDINIHSC
                KNTNISSYSTYNNMEKENINIYDKYNIHNFYTEKSISYKDENCQHITLNMIYLLNQTY
                DNICRICLNTNTNIYINFYMINILKYICYKNMEIILLNYNHIEDMKKKINQKNNTNTS
                LFKYIYSFFFFKKEENHIYDLFEDQMMNHLHKKENDKFYNYSNENTHNNIYKYISDNY
                FYDHINSSSNRCSFKNLKKQQTDDNTKHIIMGKEKYPMNKSDHEKKNNNTCGNINIEK
                DQKKDILKKIYFLKGNKLDDIQILNELYVMIYMRLLFECSLKLISIKKNIHLLEKKME
                FDKDNKIIYLNSADYMNNLRRNILKRFSKNEERENINSFASPPFLLSKNIIYFEDEIG
                RSRDNTIYNNVYDKETNKTTTNNNNNDNNDNICSNNDHICSNNNDHICSNNDHICSN
                NNNNICSNNNNNICSNNNNNICSNNNNNICSNKMLDEFCQDNKFNDYNTRKKEKRKRI
                YELAKIYTNNIFDYLKGKKEKHQNEDNTINLYYIKKKFPWIFYLKNIIKNKDTSFIEH
                NNNIVNGDIKNNNIIFKKKYNLFESSIISYFYIKDIYEYNYKLRLYYIYDNLIKKFCR
                YFLKMNEHINRKLYKMKRAFHYYIYNFDQFIINNYYHIIHKKNIHKIHIHLKQCKDKE
                IDIVKFKDLYYCMINNINNIFSYIHKVDHNECVYRIFKAYNKILLYEYNYLNEKENIY
                YKNKIKKYLTYLNNNISNDLYPYNISYNKIYNQNKYKNRKNFSHIFYSLKNDIHLLLF
                LYTQRIQNCCDIFSYIYKKYNFNEKNPFLNYLYYELHYIVYSEKKKKKKFFSFISSSP
                YSYDTMVNSFTFSYFFFSLSYLLFILFYHPDMYASYIFFKTLTYSGLPTYYYSLYNNI
                MVVCGPKT"
gene            join(7669. .7717,7800. .7829,7912. .7940,8064. .8106,
                8302. .8368,8568. .8641,8713. .8812,8924. .9006,9122. .9258,
                9369. .9505,9613. .9838)
                /gene="PFC0581w"
CDS             join(7669. .7717,7800. .7829,7912. .7940,8064. .8106,
                8302. .8368,8568. .8641,8713. .8812,8924. .9006,9122. .9258,
                9369. .9505,9613. .9838)
                /gene="PFC0581w"
                /note="PFC0581w, hypothetical protein, len: 324 aa,
                revised: new gene prediction, splicing prediction very
                tentative"
                /codon_start=1
                /product="hypothetical protein, PFC0581w"
                /protein_id="CAB90285.1"
                /db_xref="GI:7711066"
                /translation="MKKKKKHYFSIKFVNVDKNKYVLCSKDYIRIINYMIGLHIFRLL
                QKIFHIYFANEQIASSFFESHKNYRVTKEDIIDGIEKCWFNITDYLISESIKQDNDFS
                NDIKTTVTAMKNKMDQLLTTSYSNKKIDTVNASFQWAQSPEYIFLNIKFSHRWSSPGA
                LKVKDEEIVSKKNNFSFSALSNDSNSVTKKYIVDLTLLDNIIESETKYNFASVGKVVV
                TLKKEKKKIWNRLLLSKEKYPNMQVWWDMKEKRIHIITFVTINLFFLLSLSHRYHDSV
                QNFLKEEKNNSDKLQDDIDEDEEKYFDEEILREAKKKSEEYDKDDEEL"
misc_feature    7790. .7799
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 2 of
                PFC0581w"
misc_feature    7830. .7835
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 2 of
                PFC0581w"
misc_feature    7901. .7911
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 3 of
                PFC0581w"
misc_feature    7941. .7947
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 3 of
                PFC0581w"
misc_feature    8055. .8063
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 4 of
                PFC0581w"
misc_feature    8107. .8112
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 4 of
                PFC0581w"
misc_feature    8293. .8301
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 5 of
                PFC0581w"
misc_feature    8369. .8374
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 5 of
                PFC0581w"
misc_feature    8559. .8567
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 6 of
                PFC0581w"
misc_feature    8642. .8647
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 6 of
                PFC0581w"
misc_feature    8704. .8712
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 7 of
                PFC0581w"
misc_feature    8813. .8818
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 7 of
                PFC0581w"
misc_feature    8913. .8923
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 8 of
                PFC0581w"
misc_feature    9007. .9012
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 8 of
                PFC0581w"
misc_feature    9112. .9121
                /gene="PFC0581w"
                /note="potential splice acceptor sequence for exon 9 of
                PFC0581w"
misc_feature    9259. .9264
                /gene="PFC0581w"
                /note="potential splice donor sequence for exon 9 of
                PFC0581w"
misc_feature    9361. .9368
                /gene="PFC0581w"

 Query Match             5.3%;  Score 93.6;  DB 3;  Length 86827;
 Best Local Similarity   45.1%;  Pred. No. 7.3e-05;
 Matches  478;  Conservative    0;  Mismatches  569;  Indels   13;  Gaps    3;

Qy     225 tagaattacaaaaaataataaaaatataaataaaaatgatagaaaatcaccaagacaaat 284
           || || ||  ||| | |    ||  | | |   |  |  || || |     || |  ||
Db   41295 TATAAATAATAAATACACATCAATAAAATAATCATTTATTAAAATAAATATAATAATAAA 41236

Qy     285 tttagaggagtataaaaaaaggaaacaaggtattatagcaggatactatggttcatggaa 344
           ||||      ||||| | ||  ||| |   ||  |||  |  | | ||   ||||| ||
Db   41235 TTTAATTTTATATAATATAATTAAATATTATAAAATACAATTAAAATAAATTTCATAAAA 41176

Qy     345 cagtcaaggtgatagagcaaaacatatgattgattcaaacccaatggtgtcaattttata 404
           | | ||    | | |  | ||  ||   |  ||  ||    |||    | || | ||||
Db   41175 TAATTAAATATACATATAATAAATTAAATTAAATATAATATTAAT--AATTAAGTGTATA 41118

Qy     405 tattgcatttgctcgcattaatatgttatatgatgtatctagaccatttaatggaagaca 464
             | ||  |   |      ||| || ||| ||  | ||          |  |||  |   ||
Db   41117 ATATATATATAATAAATATAAAATATTAAATATTATAAAATATAATTAAAATAAATTTCA 41058

Qy     465 aagattcctattaagaaaacacggcttagaatatgaaacctatggtatgatgcttaatga 524
           | | |  |   |  | | |    | | ||| | |||   ||   ||     ||||| |
Db   41057 TAAAATAATTAAATACATATAATAATAATAATTTTAAATAAATATAATA----TTAATAA 41002

Qy     525 aattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtggaga 584
             |  ||  | | |  ||  |  | ||  | | |   |   |  |   |  |     | |
Db   41001 TTAACTATATAATATATATATAAATAAAAATTAATTATTATATAATTAGATTTAATTATA 40942

Qy     585 aacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagcttgt 644
             || || |  | |  ||| | | |||   ||  |   |||| |  |  |  || |
Db   40941 TTATTAAATTAATTAAAATAAAATAAATATTATATAATTAATTAAATTTATATATATTAT 40882

Qy     645 taatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttacaa 704
           ||| | ||  | |   ||       |||          ||  | |    | ||| || ||
Db   40881 AAATTAATTAAAATATAATAAATATTATAAATTAATTAAAATATAATAAATGTTATATAA 40822

Qy     705 cttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaac 764
           | || | ||||| ||| || || | |||| | |||||  | ||| |   |||  | ||
Db   40821 TTAAATTTAATTATATTATTAAATTAATTAAAATAAATATAATAAATATATAATTATAAT 40762

Qy     765 tattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgt 824
```

```
Db  40761 TAAATTTAATTATATTATTAAATTAATTAAAATATAATAAATATTATATAATTAAATTTA 40702

Qy    825 ttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtc 884
Db  40701 AATATATTATTTAATTAAAATAAAATAAATGTAATAAATATTATATAATTAAATTTAAAT 40642

Qy    885 tgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaac 944
Db  40641 ATATTATTTAATTAATTAAAATAAAATAAATATAATAAAATAATTATATATAATAATAAT 40582

Qy    945 ttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaacata 1004
Db  40581 TAAATTAAATTAAATTAAATAAATATTAATTAAAATAAAATACTTTAAATATAAATAATT 40522

Qy   1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
Db  40521 AAATTAAATAAATATTAATTAAAATAAAATACTTAATAAGTATATAATAATTAAATTAAA 40462

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
Db  40461 TATATTTAATTAAAATTATATTATTAAATAATTATTTAATTTTAATATTAATATTAATTA 40402

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
Db  40401 ATTAAAGTTAAAAAATATAAAATAATTAAATT-------AAATATATTAATAAATAATAA 40349

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
Db  40348 AATATATTATATTATAATATAATATATAATTATATATATTAAAATAAAATAAAATTAATA 40289

Qy   1245 accaactggatcattcgatgtagatatttttcttacaaat 1284
Db  40288 ATATATTATATTATAATAATTATATATAATAATAATTAAT 40249


RESULT  14
AC006281
LOCUS       AC006281   199551 bp    DNA             HTG       12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AC006281
VERSION     AC006281.8  GI:9797738
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 199551)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199551)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810457.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    41526: contig of 41526 bp in length
            *    41527    41726: gap of unknown length
            *    41727   199551: contig of 157825 bp in length.
FEATURES            Location/Qualifiers
     source         1. .199551
                    /organism="Plasmodium falciparum"
                    /db_xref="taxon:5833"
                    /chromosome="12"
                    /clone="PFYACB8-95"
                    /clone="3D7"
BASE COUNT    83197 a  17543 c  19759 g  78852 t    200 others
ORIGIN


  Query Match             5.3%;  Score 93.6;  DB 2;  Length 199551;
  Best Local Similarity   43.7%;  Pred. No. 6.5e-05;
  Matches  569;  Conservative    0;  Mismatches  719;  Indels   14;  Gaps    3;

Qy      20 tatttttaattatagtatccatcttgtattctgcaaattccagaaccttgaaaggaaaaa 79
Db  177411 TATGTATTACTATCATTCTCTTCTTCTGATCCGTCATTATAAGACATTTTAAAAAAAAAA 177470

Qy      80 ataatataaataattcattgggaataatacgggaaaataaaaataaaactcatcaaacgg 139
Db  177471 AAAAAAAAAAAAAAAAATATATATAAATATATATATATAAAATTATAAATATATAAAATA 177530

Qy     140 aaatacatgagtctttttcacatcttaaatcgaataatagtaattttgtagaatatggat 199
Db  177531 AAATAAAAACAAATAAAAAAAAAAATTGTATGGACATATATATATTATATGTATATTTCAT 177590

Qy     200 cttattgcggagatgggtgtaactctagaattacaaaaaatataataaaaatataaataaaa 259
Db  177591 ATATATGTTGAATAAAATACAAATGTACTACTAAATAAAAGAATAAATAAATATTAATAA 177650

Qy     260 atgatagaaaatcaccaa--gacaaatttagaggagtataaaaaaaggaaacaaggtat 317
Db  177651 CTGAATTATTATTAAAAAATAAACAACTTAAAAAAAAAAAAAAAAAAAAAAATACATCCTT 177710

Qy     318 tatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacatatgattga 377
Db  177711 ATCATTTCAATAATATATATAATATGAATATTATATATATATATATATATATATATAA 177770

Qy     378 ttcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatgttatatga 437
Db  177771 TCGTAACAATTTTACAAATACATTTTTATATACATATTATTGATATGTTTTTTTTAATAA 177830

Qy     438 tgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggcttagaata 497
Db  177831 GTTAATTTCTTCTTTCATTTTCTTGACATATATGAATTTCATTAATATAATGTGATAGTG 177890

Qy     498 tgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagatgt 557
Db  177891 TAGTAAATATAATATCGAGTGTTCATATTAATAGAGAATAAAATAAAAAAAT-AATATGA 177949

Qy     558 aattattcttttatccttaggtggagaaacctatatgatagatatagaaaaagaaattga 617
Db  177950 AATGAATATATAAATATTATATATCCATACATATATATATAATATGTATGCATATATTTA 178009

Qy     618 ttatgtggataaaatattgaagcttgttaatgattttgatttagatggtgtagatattga 677
Db  178010 TATATATTATATATATATATATATATATAAATATATATATATATATATTTTTTTTTTTATGT 178069

Qy     678 ctgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaaattattatat 737
Db  178070 ATTTTATTCCTCTTTAAAAATGTGCATATTTTATAAAATATATTATAATTGTATTTTTTT 178129

Qy     738 taaattaattaacttgttaagaaaaactattccggaagaaaagttaatttcaatttctgg 797
Db  178130 TTTTTTTTTATTCTTTGGAAGAACAATAATTTACTTCATAGAAAAGAAAAAAAAAGCACA 178189

Qy     798 ttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttctgtaaagatgaaga 857
Db  178190 TAAAATTAAAAATATAATAATTCCTTGTGCTAAAGATAAAATTACATTATATTCATTTAT 178249

Qy     858 atctccatataacactaaatttttgtctgaacaaatagaaacaaataaagaattacatag 917
Db  178250 ATATTACTTTTATAATATATGAATATCTTTATACAATAAAAAAAAAAAAAAAAGGAATCC 178309

Qy     918 ggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaaggagaa 977
```

```
Db 178310 ACTACTTTTCATTATGATAGGAAAGAGGAATATTCTTCTCTTTTTATTGTTGAATTAATC 178369

Qy    978 aatagatcttgtatt----------tattcaaacatacaatttagaaactacaaatcca 1026

Db 178370 ATTTGTTTGTATGTTAAATGGTAAAATAATCAAACATAAATATAAATAAATAAATATAAA 178429

Qy   1027 gatataatggtagatatgtacttatcccatttatattttggtttaaaatataacatcaca 1086

Db 178430 TATATATATATATATATATATATTGTGATGGCAAATGGATAGAGGTAATAAAAATATA 178489

Qy   1087 atcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaataaagaa 1146

Db 178490 ATTGACCTGGAATTAATTTCTTATTTATTATATGATTAATTATATGGTTCTCATAAAGTA 178549

Qy   1147 ttattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagggcagat 1206

Db 178550 CTTCTTATTTTATAAACAAAAAAAAAGAAAGAAAAAAAAAATAAATAAAAAAAATAAATA 178609

Qy   1207 ggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattcgatgta 1266

Db 178610 AATAAATAAATAAATAAATATATATATATATATATATTGAAAAAGACAATGGATATAA 178669

Qy   1267 gatatttttcttacaaatatttggaaacatttaaatcctgaa 1308

Db 178670 TATATATTAGGTACAAATTTGAAAGAAACTATTAAAATTAAA 178711


RESULT  15
AF396436/c
LOCUS       AF396436    47577 bp    DNA             INV       25-JUL-2001
DEFINITION  Tetrahymena thermophila strain SB210 mitochondrion, complete
            genome.
ACCESSION   AF396436
VERSION     AF396436.1  GI:15011465
KEYWORDS    .
SOURCE      Tetrahymena thermophila.
  ORGANISM  Mitochondrion Tetrahymena thermophila
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 47577)
  AUTHORS   Brunk,C.F., Tran,A.B., Lee,L.C. and Li,J.
  TITLE     Complete Sequence of the Mitochondrial Genome of Tetrahymena
            thermophila and Comparison With the Mitochondrial Genome of
            Tetrahymena Pyriformis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 47577)
  AUTHORS   Brunk,C.F., Tran,A.B., Lee,L.C. and Li,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2001) Organismic Biology, Ecology and Evolution,
            University of California, Los Angeles, 621 Charles E. Young Drive
            South, Los Angeles, CA 90095, USA
FEATURES             Location/Qualifiers
     source          1. .47577
                     /organism="Tetrahymena thermophila"
                     /organelle="mitochondrion"
                     /strain="SB210"
                     /db_xref="taxon:5911"
     repeat_region   complement(1. .53)
                     /note="telomere repeat"
     tRNA            complement(65. .145)
                     /gene="trnY"
                     /product="tRNA-Tyr"
                     /anticodon=(pos:110. .112,aa:Tyr)
     gene            complement(65. .145)
                     /gene="trnY"
     misc_feature    complement(161. .2844)
                     /note="copy number 1; includes rnl_a, trnL and rnl_b"
     rRNA            complement(161. .438)
                     /gene="rnl_a_1"
                     /product="large subunit ribosomal RNA"
     gene            complement(161. .438)
                     /gene="rnl_a_1"
     tRNA            complement(448. .528)
                     /gene="trnL_1"
                     /product="tRNA-Leu"
                     /anticodon=(pos:492. .494,aa:Leu)
     gene            complement(448. .528)
                     /gene="trnL_1"
     rRNA            complement(531. .2844)
                     /gene="rnl_b_1"
                     /product="large subunit ribosomal RNA"
     gene            complement(531. .2844)
                     /gene="rnl_b_1"
     gene            complement(3023. .3325)
                     /gene="ymf57"
     CDS             complement(3023. .3325)
                     /gene="ymf57"
                     /note="CPGN"
                     /codon_start=1
                     /transl_table=4
                     /product="ymf57"
                     /protein_id="AAK77561.1"
                     /db_xref="GI:15011466"
                     /translation="MLKNKLIKFKFFRFVQSGFYVDFIFKKFSEMFIRNIFIYSSIFF
                     GEKFMIEYLTKKTIDSFIFNNNRFNFINLVESKYFLQILTLILYLFFITIFILFYI"
     gene            complement(3335. .4675)
                     /gene="ymf66"
     CDS             complement(3335. .4675)
                     /gene="ymf66"
                     /note="CPGN"
                     /codon_start=1
                     /transl_table=4
                     /product="ymf66"
                     /protein_id="AAK77562.1"
                     /db_xref="GI:15011467"
                     /translation="MGRENVLPVHNDVYEDFVFTTPYFQPESTFKSVPKLFSDILLGG
                     VEWVYTTSESVLAYDYKLWYLWSGVSNLDESFDMFFNQYWALSLSTSVFQLFYAVILD
                     RYLSVLFQNTPYTNDWFRMMLHSKETALIWLYHPELSWHINGLNQFFTYFYGGILEFV
                     YFDKSNPDMCILVHTLWIHLLILFLIFTGFVTILFSFYGNPNTEENTIDSDYLAASGT
                     VEAEKEITSIDDYLGLVFAIAYVFGVFFYVHGWTSMLSHAVLLLSCYSIIIMFLFILG
                     MPTLLLYDFGIFFLAYLKGAGKYISSVAEMMFDYTACLVFYIRILAQWIRVVLMVVTF
                     ISLSHYVSDFDITNSALIGSENQSDSMNELNTNFSMTYYILTVLPGKFIYWIYEILHT
                     FFVVCSQFVAFFAIVFWLFLFLYTFFIIEKHEDFFSKKREERKKKLKELWNLKN"
     gene            complement(4937. .6154)
                     /gene="ymf76"
     CDS             complement(4937. .6154)
                     /gene="ymf76"
                     /note="CPGN"
                     /codon_start=1
                     /transl_table=4
                     /product="ymf76"
                     /protein_id="AAK77563.1"
                     /db_xref="GI:15011468"
                     /translation="MFLVKFKIKKLRLKKKIKKFYKLINYNFSNLLNNFYHKKPNFLT
                     LYNNTNNFFLKILFYIKYINLISKTISNKKLFKFLNNPKIRNRKKFKYKYSDKIKFIL
                     NILKSKKTKIKNLLFFIKYFSVLRKRQSRIFNLARVKSRLSKRRFFKKKLKKKKIAKY
                     FFQMFKKLKFKHKKYINLINLDFYFIRNKRFFRLHRLYDIRKKYIRYLNNNRNIYKFY
                     KFRIKHNFKFIKRHIKSISKLSIKDRVHFYELSLRNIAIKLKYAFTLRNANLFTKSGF
                     IFLNGHQELNPFKYAYKGDIIELPFSKFILKLRRKMKKKMFNSMRKYKKYNWRTLKNK
                     VNPEQRRLRISRFSENTLNFKTKLTKLFQYDYRTLSYCVVLDTNFKRDLTYLNKKLIP
                     IYLLKLFNWKIIS"
     gene            complement(6189. .7019)
                     /gene="rps13"
     CDS             complement(6189. .7019)
                     /gene="rps13"
                     /codon_start=1
                     /transl_table=4
                     /product="ribosomal protein S13"
                     /protein_id="AAK77564.1"
                     /db_xref="GI:15011469"
                     /translation="MKEFFQLNGYIFTKDTEFINTKKKIFGWNDYTYKIIYNRFEKFY
                     SENLISYKDTEWKKFKSILSTFFPKSRSLKNRMDTNIFFLDFLNTYRGLRHSKGLPVR
                     GQRTWTNAWSVYRSNLNLRKFKLEIAKRIYGNMPTPSLTTIYLAEQINYIWKLQWKRE
                     WLQARSKRLKLMQNEHAMFKIDINSMSKGYIDGFDKKKEITKKKKAQAKKNTFTLGFE
                     PGFTLFYLRPNNSLSAKERNKIKIVLNEEENKVKILQKKKKVVLKKKPEKKKKNSSWE
```

```
                "
     gene       complement(7023. .7478)
                /gene="rps3"
     CDS        complement(7023. .7478)
                /gene="rps3"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S3"
                /protein_id="AAK77565.1"
                /db_xref="GI:15011470"
                /translation="MGLKSLPMLNKSGISMYWHNIWDSIKLYKKYSLSFLFLNEVINH
                FLNENLYYYCIMKIRPTDPRLKGFRGNKSININKIKKSWNMRHFYLGKILFLKYQGWV
                LVLINYYCSRRNKLYINYKSFKAFKKIAKSFRQGVTSYVYKMDKYKFKF"
     gene       complement(7483. .7779)
                /gene="rps19"
     CDS        complement(7483. .7779)
                /gene="rps19"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S19"
                /protein_id="AAK77566.1"
                /db_xref="GI:15011471"
                /translation="MGKSHWRLKYISKSIISKIIKESIGIKIKKKISVIMNKSSTIPS
                SLTDNTFYIHKGYKFRELKIKDYHVGFKFGEFILTRKPNIFPKKSKNKSKNFRR"
     gene       complement(7784. .8572)
                /gene="rpl2"
     CDS        complement(7784. .8572)
                /gene="rpl2"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein L2"
                /protein_id="AAK77567.1"
                /db_xref="GI:15011472"
                /translation="MFLKNYKPKTPTLRFKRDIFKMTPFIKKNNIFFIKNKNKSGRST
                KGQIIVRHRGANKYNKHLSINVNRNQLTNLAIITSLNYINKNSCFIGLIKYSNNCLSY
                IKLANGLFIGNYTKSLDKPLNFSFNFRNILGSFIILNLAPKHSIFSNLISFNKKKSNY
                AKSAGTFLSIIKKIKEINLFILKLPTGEKKYFSGDSKATLGRNSNFLNKFTVTGKAGI
                NRNKGWRSSVRGVAMNPVDHPHGGRTKTNQPEMSPWGWITKKNK"
     gene       complement(8576. .9049)
                /gene="ymf74"
     CDS        complement(8576. .9049)
                /gene="ymf74"
                /note="CPGN"
                /codon_start=1
                /transl_table=4
                /product="ymf74"
                /protein_id="AAK77568.1"
                /db_xref="GI:15011473"
                /translation="MKFKREFKFLIKKKNFKFKKFKILLKIYYSIKNLINFYKIIKLN
                NFKIKSTLLINNNYYFNYITNGLDLKYDNTFQNFELNTLSIKNYKNKNLIISNNNQLD
                IIKFQKFLFIIDNKYVNSLICDNLFDFFFISIILTNSLILEFYKNIILINLIKIN"
     gene       complement(9062. .9550)
                /gene="nad10"
     CDS        complement(9062. .9550)
                /gene="nad10"
                /EC_number="1.6.5.3"
                /codon_start=1
                /transl_table=4
                /product="NADH dehydrogenase subunit 10"
                /protein_id="AAK77569.1"
                /db_xref="GI:15011474"
                /translation="MRVDYIVTPINAVVSWARQGSFWPLTFGLACCAVEMMHATVSRY
                DFDRFGVIFRATPRQADLILVAGTLTNKMAPGLRRLYEQTLDPKWVLSMGSCANGGGY
                YHYSYSVVRGCNRIIPVDIYVPGCPPTAEALMFGVLQLQKSLYRQINEGEKKYNKKDL
                TY"
     gene       complement(9567. .9968)
                /gene="rps12"
     CDS        complement(9567. .9968)
                /gene="rps12"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S12"
                /protein_id="AAK77570.1"
                /db_xref="GI:15011475"
                /translation="MGSLCINKKKGNLHRKLRARILGINPQKKGVVMKARIVTPRKPN
                SARRPVAKVILVNKKRLTAHIPGIGHNIRRHSSVLVRGGGCRDLPGVRYTCVRGVYDF
                SKVLNRFNRRSRYGIPKPDEEKKKLRRKFRV"
     gene       complement(10019. .10555)
                /gene="nad2"
     CDS        complement(10019. .10555)
                /gene="nad2"
                /EC_number="1.6.5.3"
                /codon_start=1
                /transl_table=4
```

```
  Query Match             5.2%;  Score 92;  DB 3;  Length 47577;
  Best Local Similarity   44.5%;  Pred. No. 0.00013;
  Matches  679;  Conservative    0;  Mismatches  825;  Indels   22;  Gaps    7;

Qy      6 ttttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccagaac 65
          ||| ||  |   | |  | ||| |||| ||   ||    | || |  || ||  | |
Db  26633 TTTAAATTTTATATTTCTATTATTTATATTAGAAATAAATTTTTTTTTAACTTTAATATA 26574

Qy     66 cttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaataa 125
          ||   |   |  ||||| ||           | | |  ||   |  | |    | ||
Db  26573 CTGTTATATATTTAATAAAATTTTATGAGATATAGTAGCAACTTGATATATAGGAGATTT 26514

Qy    126 aactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaatttt 185
            |   ||   |   |||  ||      | |  | || | | |   || |||  || | |
Db  26513 TATAAATGGTAAATTAATTTATATAGAATCTAAATATATAAGAAAAAAAAATTATATTGT 26454

Qy    186 tgta-----gaatatggatcttattgcggagatgggtgtaactctagaattacaaa-aaa 239
            ||      | ||| ||| | |||   ||  ||  |     | || ||||| ||| |||
Db  26453 ATTAATTACAATTATAGATTTCATTTATGATGTGATTTATGATATAAAATTAAAAATAAA 26394

Qy    240 taataaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataa 299
          | ||| ||| || |   ||||   ||     |||   | | |||    | |     | |
Db  26393 TGATATAAAAATTATCGAAAAATTTAATCTCTCAAATAAAAAAAAAAAAAATTTAAAAGA 26334

Qy    300 aaaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatag 359
          |||||  ||  |   ||  |  |           ||  ||  |      |  | |||
Db  26333 AAAAATTAATAACATAATATTTATATTTATTAGAATTTTGTGATATCTATGATTTAATAT 26274

Qy    360 agcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcg 419
            | |   | | | |  |   ||     |    | ||| |||  ||      |   |
Db  26273 TAAATATGTTTTAAATTTTATTAATTATTTATATTTAATATTACTTAGATATCTATATAT 26214

Qy    420 cattaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaag 479
            |||||   || || |      || |  |  |||     ||   |||||      ||
Db  26213 ATATAATAAAATAAATTACAATAATAAATTAAGTAAATCTTGAATTAGATTGAATACAAA 26154

Qy    480 aaaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcag 539
          ||||     | | |||||  |    ||  ||  |    || ||||||||  |   | |
Db  26153 AAAATTAATTTCAAAATATATATTAAATTTTAATAAAAGTATTGAAATTAAAAAAACCTT 26094

Qy    540 aaaagtacgt-ccagatgtaattattcttttatccttaggtggagaaacctatatgatag 598
          | | |||  |   | |   | | ||| | |     ||    |         ||  | ||
Db  26093 ATATGTAAATAATATAAAAATTAATTATCTAGAAATTTTAAGTTTCTTAATAATTAATGA 26034

Qy    599 atatagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatt 658
          | | | |||   |||   |||    | ||| ||   |  |  |  ||| |     |  ||
Db  26033 AAAAAAAAACGAAAAAATATTTGAAGAATATAAATCTATAAGTATTTATTTTAAATTTTT 25974

Qy    659 tagatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaa 718
          || |   | || |   | | |        |  |  || || |  ||    |   | ||
Db  25973 TAAAAATTTTAAAATATTATTTATTTTAATTTTAAAATTATTAAAAAAGATGTATTTTT 25914

Qy    719 atttttcaaattattatatt-----aaattaattaacttgttaagaaaaactattccgga 773
          |||  | |||| ||||||||     |||||||  | | || |  | |||| ||||  ||
Db  25913 ATTAATAAAATGATTATATTTATTAAAATTAACTTATTTTTATATAAAAAGTATTCTAGA 25854

Qy    774 agaaaagttaatttcaattt------ctggttcatcaaatgctgcattatcatgcgtttc 827
          |  ||| ||||  | |||||       |  ||     |||  |    ||        |
Db  25853 AATAAAATTAAACTTAATTTTATATAATATTTATAATAATAATTATATAAATAAATATAT 25794

Qy    828 aggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtctga 887
```

```
Db  25793 ATATATATTAACATATAATACAAAATTAAAATATAATAATCTAATTAAATATAAAAATAA 25734

Qy    888 acaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaacttt 947
Db  25733 TAATATATACTTATATAATAATAATTATAGTAAATATAATAATAAATTAAAAATAAATAT 25674

Qy    948 tattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaacatacaa 1007
Db  25673 AAGTAAAATAATAAAAAATTTTAAAATAAATATAAATTATAATAATATATATATATAA 25614

Qy   1008 tttagaaactacaaatccagatataat---ggtagatatgtacttatcccatttatattt 1064
Db  25613 ATTAAAAAATTTATATAATTATATAATTAATATAAATTGGAAAGAAATTATATTAGAATT 25554

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
Db  25553 AAAAATTTTATATATAAATAGTAAAAGGCCATTTAAACAACTTAAATATGATAAAAGTTT 25494

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
Db  25493 ATATATTATAAAAATTTTAGATGGATTAGAATATGCTATGTTAACAAATTTTAAATCATA 25434

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
Db  25433 TAAAAATAAAAATATAATAAAATATCATAAAAAAAAAATATAAATTTAAAAAAATACCA-A 25375

Qy   1245 accaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaatcc 1304
Db  25374 GTAAAATAGAAAAATATATAAATACAATAAAACCTATAATAAAAAAAAATATTATAGAAA 25315

Qy   1305 tgaagtacaaactccaaaagaccttactataactgaaaaccctgaagactgtagcacaat 1364
Db  25314 TATTAGAAAAAATAAAAATTAAAATAAATAACTATAAAGAAGTAATAAGTTTAATTAATA 25255

Qy   1365 agatgaatatgttccaggactcgttattccaaccatagggatatattacaaacacaatga 1424
Db  25254 TAAATAAAAAAAATAAAATTACTTATATAAATTTAGATATATATAGTTTAATTAAAATTA 25195

Qy   1425 tgctatatggaaaactagatcttattcaattcatgcacctggtgtagacagatatgaatg 1484
Db  25194 ATTTATACATAATTTTTAAAAATATTATACCAAGCCACCTTAATATTATACTAATTAAAT 25135

Qy   1485 ggacttggtcaaagtatgctatgaaa 1510
Db  25134 TGAATAATCCAATGTTATTAATAAAA 25109


Search completed: March 21, 2002, 15:52:35
Job time: 5086 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:29:29 ; Search time 174.24 Seconds
(without alignments)
8679.532 Million cell updates/sec

Title: US-09-579-383-2
Perfect score: 1764
Sequence: 1 atgaattttaaaatatcaat.....tggatttaccattacaataa 1764

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1764	100.0	1764	22	AAC89669	P. gallinaceum chi
	2	1764	100.0	2500	22	AAC89670	P. gallinaceum chi
c	3	219	12.4	936	22	AAF58252	Oligonucleotide D1
c	4	219	12.4	936	22	AAF58254	Oligonucleotide D1
c	5	219	12.4	936	22	AAF58257	Oligonucleotide D1
c	6	219	12.4	936	22	AAF58259	Oligonucleotide D2
c	7	219	12.4	936	22	AAF58262	Oligonucleotide D2
c	8	219	12.4	938	22	AAF58255	Oligonucleotide D1
	9	217.2	12.3	936	22	AAF58252	Oligonucleotide D1
	10	217.2	12.3	936	22	AAF58254	Oligonucleotide D1
	11	217.2	12.3	936	22	AAF58257	Oligonucleotide D1
	12	217.2	12.3	936	22	AAF58259	Oligonucleotide D2
	13	217.2	12.3	936	22	AAF58262	Oligonucleotide D2
	14	217.2	12.3	938	22	AAF58255	Oligonucleotide D1
	15	104.2	5.9	1137	22	AAC89668	P. falciparum chit
	16	73.2	4.1	244	22	AAF58238	Oligonucleotide D1
c	17	71	4.0	244	22	AAF58238	Oligonucleotide D1
	18	66.4	3.8	11922	21	AAA70187	Plasmodium falcipa
	19	63.8	3.6	7458	21	AAA70106	Plasmodium falcipa
	20	63.4	3.6	3101	11	AAQ02047	Sequence encoding
	21	62.6	3.5	6621	21	AAA70188	Plasmodium falcipa
	22	62.2	3.5	19124	18	AAT72882	Plasmodium var-7 g
	23	62.2	3.5	19124	21	AAZ98287	Plasmodium var-7 p
	24	61.2	3.5	3642	21	AAA70180	Plasmodium falcipa
	25	61	3.5	3399	17	AAT05868	Chicken leucocytoz
	26	59.6	3.4	2503	15	AAQ53480	pNPX30 xylanase cD
	27	59.6	3.4	6042	21	AAA70199	Plasmodium falcipa
	28	59.6	3.4	9789	17	AAT41852	cDNA encoding Plas
c	29	58.6	3.3	4838	20	AAV70891	Plasmid pTK-esp-la
	30	58.4	3.3	5940	21	AAA70105	Plasmodium falcipa
	31	58.2	3.3	5145	21	AAA70209	Plasmodium falcipa
c	32	58	3.3	665	21	AAZ31996	Human apolipoprote
	33	58	3.3	1671	13	AAQ24134	50 kD subunit of S
	34	57.8	3.3	4197	21	AAA70227	Plasmodium falcipa
	35	57.8	3.3	6033	21	AAA70152	Plasmodium falcipa
	36	57.8	3.3	6644	20	AAX33181	Base sequence of t
	37	57.8	3.3	7372	20	AAX33182	Base sequence of t
	38	57.8	3.3	7797	20	AAX33180	Cowpox virus bsr f
	39	57.8	3.3	7996	20	AAX33184	Base sequence of t
	40	57.6	3.3	765	21	AAC59300	Human secreted pro
	41	57	3.2	612	22	AAH71471	Human cervical can
c	42	57	3.2	660	15	AAQ66819	AmEPV thymidine-ki
c	43	57	3.2	1511	13	AAQ28302	AmEPV tk DNA. Ams
c	44	57	3.2	1511	15	AAQ66798	AmEPV thymidine-ki
c	45	57	3.2	1511	19	AAV14508	AmEPV entomopoxvir

ALIGNMENTS

RESULT 1
AAC89669
ID AAC89669 standard; DNA; 1764 BP.
XX
AC AAC89669;
XX
DT 13-MAR-2001 (first entry)
XX
DE P. gallinaceum chitinase coding sequence SEQ ID NO: 2.
XX
KW Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KW veterinary disease; arthropod pest; ds.
XX
OS Plasmodium gallinaceum.
XX
PN WO200073488-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14536.
XX
PR 28-MAY-1999; 99US-0136508.
PR 03-FEB-2000; 2000US-0180051.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Vinetz JM;
XX
DR WPI; 2001-061553/07.
DR P-PSDB; AAB49986.
XX
PT New nucleic acid encoding a Plasmodium species chitinase is useful for
PT preventing transmission of malaria by mosquito feeding on subject that
PT may harbor Plasmodium species organisms -

```
XX
PS   Claim 4; Page 123-124; 137pp; English.
XX
CC   The present invention provides the protein and coding sequences of the
CC   Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC   organisms are the cause of malaria in humans. The sequences are useful
CC   in the prevention and treatment of malaria, fungal diseases, parasitic
CC   diseases and veterinary diseases, in preventing the transmission of
CC   malaria and in the control of arthropod pests in agriculture.
XX
SQ   Sequence 1764 BP; 717 A; 227 C; 294 G; 526 T; 0 other;


  Query Match             100.0%;  Score 1764;  DB 22;  Length 1764;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-287;
  Matches 1764;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaatttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgaatttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattcc 60

Qy     61 agaaccttgaaaggaaaaaatataataaataattcattgggaataatacgggaaaataaa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 agaaccttgaaaggaaaaaatataataaataattcattgggaataatacgggaaaataaa 120

Qy    121 aataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagt 180

Qy    181 aattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaat 240

Qy    241 aataaaaatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aataaaaatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaa 300

Qy    301 aaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgataga 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 aaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgataga 360

Qy    361 gcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgc 420

Qy    421 attaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaaga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 attaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaaga 480

Qy    481 aaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcaga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 aaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcaga 540

Qy    541 aaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 aaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagat 600

Qy    601 atagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgattta 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 atagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgattta 660

Qy    661 gatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 720

Qy    721 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 780

Qy    781 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 840

Qy    841 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 900

Qy    901 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960

Qy    961 aatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 aatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1020

Qy   1021 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1080

Qy   1081 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1140

Qy   1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200

Qy   1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1260

Qy   1261 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1320

Qy   1321 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1380

Qy   1381 ggactcgttattccaaccatagggatatattacaaacacaatgatgctatatggaaaact 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ggactcgttattccaaccatagggatatattacaaacacaatgatgctatatggaaaact 1440

Qy   1441 agatcttattcaattcatgcacctggtgtagacagatatgaatgggacttggtcaaagta 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 agatcttattcaattcatgcacctggtgtagacagatatgaatgggacttggtcaaagta 1500

Qy   1501 tgctatgaaaaaatatgcgatgggaaagcagcccattattataacactgactataaagaa 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 tgctatgaaaaaatatgcgatgggaaagcagcccattattataacactgactataaagaa 1560

Qy   1561 agctctattattatatggaaaggggaaccatatttaattaaatggtggcaacaaggacct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 agctctattattatatggaaaggggaaccatatttaattaaatggtggcaacaaggacct 1620

Qy   1621 ccggaaggtcaggcactagagtcatacacaaaactagatgcatccaaatgtccagggata 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ccggaaggtcaggcactagagtcatacacaaaactagatgcatccaaatgtccagggata 1680

Qy   1681 gaagaatggaataaaaaatatccacataaaccactagaagtagaggaacaatatgaacaa 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 gaagaatggaataaaaaatatccacataaaccactagaagtagaggaacaatatgaacaa 1740

Qy   1741 gaagtggatttaccattacaataa 1764
          |||||||||||||||||||||||
Db   1741 gaagtggatttaccattacaataa 1764


RESULT   2
AAC89670
ID   AAC89670 standard; DNA; 2500 BP.
XX
AC   AAC89670;
XX
DT   13-MAR-2001  (first entry)
XX
```

DE P. gallinaceum chitinase coding sequence SEQ ID NO: 5.
XX
KW Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KW veterinary disease; arthropod pest; ds.
XX
OS Plasmodium gallinaceum.
XX
PN WO200073488-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14536.
XX
PR 28-MAY-1999; 99US-0136508.
PR 03-FEB-2000; 2000US-0180051.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Vinetz JM;
XX
DR WPI; 2001-061553/07.
XX
PT New nucleic acid encoding a Plasmodium species chitinase is useful for
PT preventing transmission of malaria by mosquito feeding on subject that
PT may harbor Plasmodium species organisms -
XX
PS Disclosure; Page 127-128; 137pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC organisms are the cause of malaria in humans. The sequences are useful
CC in the prevention and treatment of malaria, fungal diseases, parasitic
CC diseases and veterinary diseases, in preventing the transmission of
CC malaria and in the control of arthropod pests in agriculture.
XX
SQ Sequence 2500 BP; 993 A; 283 C; 365 G; 858 T; 1 other;

Query Match 100.0%; Score 1764; DB 22; Length 2500;
Best Local Similarity 100.0%; Pred. No. 2.1e-287;
Matches 1764; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgaattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 atgaattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattcc 335

Qy     61 agaaccttgaaaggaaaaaatatataaataattcattgggaataatacgggaaaataaa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336 agaaccttgaaaggaaaaaatatataaataattcattgggaataatacgggaaaataaa 395

Qy    121 aataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    396 aataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagt 455

Qy    181 aattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 aattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaat 515

Qy    241 aataaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 aataaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaa 575

Qy    301 aaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgataga 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 aaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgataga 635

Qy    361 gcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    636 gcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgc 695

Qy    421 attaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaaga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    696 attaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaaga 755

Qy    481 aaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcaga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    756 aaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcaga 815

Qy    541 aaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    816 aaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagat 875

Qy    601 atagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgattta 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    876 atagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgattta 935

Qy    661 gatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    936 gatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaat 995

Qy    721 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    996 ttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaag 1055

Qy    781 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1056 ttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatct 1115

Qy    841 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1116 ttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaaca 1175

Qy    901 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1176 aataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 1235

Qy    961 aatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1236 aatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaactaca 1295

Qy   1021 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1296 aatccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataac 1355

Qy   1081 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1356 atcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaat 1415

Qy   1141 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1416 aaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagg 1475

Qy   1201 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1476 gcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattc 1535

Qy   1261 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1536 gatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaactcca 1595

Qy   1321 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1596 aaagaccttactataactgaaaaccctgaagactgtagcacaatagatgaatatgttcca 1655

Qy   1381 ggactcgttattccaaccatagggatatattacaaacacaatgatgctatatggaaaact 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1656 ggactcgttattccaaccatagggatatattacaaacacaatgatgctatatggaaaact 1715

Qy   1441 agatcttattcaattcatgcacctggtgtagacagatatgaatgggacttggtcaaagta 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1716 agatcttattcaattcatgcacctggtgtagacagatatgaatgggacttggtcaaagta 1775

Qy   1501 tgctatgaaaaaatatgcgatgggaaagcagcccattattataacactgactataaagaa 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1776 tgctatgaaaaaatatgcgatgggaaagcagcccattattataacactgactataaagaa 1835
```

```
Qy   1561 agctctattattatatggaaaggggaaccatatttaattaaatggtggcaacaaggacct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1836 agctctattattatatggaaaggggaaccatatttaattaaatggtggcaacaaggacct 1895

Qy   1621 ccggaaggtcaggcactagagtcatacacaaaactagatgcatccaaatgtccagggata 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1896 ccggaaggtcaggcactagagtcatacacaaaactagatgcatccaaatgtccagggata 1955

Qy   1681 gaagaatggaataaaaaatatccacataaaccactagaagtagaggaacaatatgaacaa 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1956 gaagaatggaataaaaaatatccacataaaccactagaagtagaggaacaatatgaacaa 2015

Qy   1741 gaagtggatttaccattacaataa 1764
          |||||||||||||||||||||||
Db   2016 gaagtggatttaccattacaataa 2039
```

RESULT 3
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 12.4%; Score 219; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.7e-28;
Matches 6; Conservative 569; Mismatches 214; Indels 0; Gaps 0;

```
Qy    523 gaaattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtgga 582
          | ::::: :  ::: : :::: ::  :  : :: :::::::: :::::: :::  :  :
Db    789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730

Qy    583 gaaacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagctt 642
           :::  ::::: ::: ::::: ::::: ::::: :::::|:  :::::::::: ::  ::
Db    729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWW 670

Qy    643 gttaatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttac 702
          :::::: ::::: ::::: ::  : :: ::::: : :   ::  : ::   :: :::::
Db    669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610

Qy    703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
          :: :::::: ::::::: :::: :::::::::::::::::::::::: :: :::: ::::
Db    609 WWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550

Qy    763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
          : ::::    :: :::: ::::::: ::::: :  :: :: :::  :  ::::: ::
Db    549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWW 490

Qy    823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
          ::: :  : ::  :: ::: : :::: :: :: ::: :  :::::: : :::::::::
Db    489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430

Qy    883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
          : : :: ::::: ::: :::::::: ::::: ::   ||:  :   :: :::: :  :  :
Db    429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWW 370

Qy    943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
          : :::::::::::::::::::: :  :::  : :::::: :: :: :::::::: ::: :
Db    369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310

Qy   1003 tacaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatat 1062
          :: :::::: ::: :  ::::  : :::::::  :: :::: :: ::::   ::::::::
Db    309 WWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250

Qy   1063 tttggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggt 1122
          :::  :::::::::::: :: : ::: ::::::  ::::: :::: :: ::::|: : :
Db    249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWW 190

Qy   1123 ggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaa 1182
            ::::: :    :::::::: :::::::: ::::  ::  ::::: :::: :: :::::
Db    189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130

Qy   1183 aatcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaa 1242
          ::: ::::::::::::    : ::  :::: | ::::   :::::::::: ::: :: ::
Db    129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWW 70

Qy   1243 ttaccaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaat 1302
          :::  :: :  :: :::  :: :: :::::::: ::: ::::::::  ::: ::::::::
Db     69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10

Qy   1303 cctgaagta 1311
            : :: ::
Db      9 WWWWWWWWW 1
```

RESULT 4
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 12.4%; Score 219; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.7e-28;
Matches 6; Conservative 569; Mismatches 214; Indels 0; Gaps 0;

```
Qy    523 gaaattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtgga 582
Db    789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730

Qy    583 gaaacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagctt 642
Db    729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWW 670

Qy    643 gttaatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttac 702
Db    669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610

Qy    703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
Db    609 WWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550

Qy    763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
Db    549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWW 490

Qy    823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
Db    489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430

Qy    883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
Db    429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWW 370

Qy    943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
Db    369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310

Qy   1003 tacaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatat 1062
Db    309 WWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250

Qy   1063 tttggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggt 1122
Db    249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWW 190

Qy   1123 ggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaa 1182
Db    189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130

Qy   1183 aatcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaa 1242
Db    129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWW 70

Qy   1243 ttaccaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaat 1302
Db     69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10

Qy   1303 cctgaagta 1311
Db      9 WWWWWWWWW 1
```

RESULT 5
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 12.4%; Score 219; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.7e-28;
Matches 6; Conservative 569; Mismatches 214; Indels 0; Gaps 0;

```
Qy    523 gaaattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtgga 582
Db    789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730

Qy    583 gaaacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagctt 642
Db    729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWW 670

Qy    643 gttaatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttac 702
Db    669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610
```

```
Qy    703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
Db    609 WWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550

Qy    763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
Db    549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWW 490

Qy    823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
Db    489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430

Qy    883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
Db    429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWW 370

Qy    943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
Db    369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310

Qy   1003 tacaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatat 1062
Db    309 WWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250

Qy   1063 tttggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggt 1122
Db    249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWW 190

Qy   1123 ggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaa 1182
Db    189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130

Qy   1183 aatcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaa 1242
Db    129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 70

Qy   1243 ttaccaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaat 1302
Db     69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10

Qy   1303 cctgaagta 1311
Db      9 WWWWWWWWW 1
```

RESULT 6
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;

XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 12.4%; Score 219; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.7e-28;
Matches 6; Conservative 569; Mismatches 214; Indels 0; Gaps 0;

```
Qy    523 gaaattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtgga 582
Db    789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730

Qy    583 gaaacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagctt 642
Db    729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWW 670

Qy    643 gttaatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttac 702
Db    669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610

Qy    703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
Db    609 WWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550

Qy    763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
Db    549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWW 490

Qy    823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
Db    489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430

Qy    883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
Db    429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWW 370

Qy    943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
Db    369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310

Qy   1003 tacaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatat 1062
Db    309 WWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250

Qy   1063 tttggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggt 1122
Db    249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWW 190

Qy   1123 ggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaa 1182
Db    189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130

Qy   1183 aatcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaa 1242
Db    129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWW 70

Qy   1243 ttaccaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaat 1302
```

```
Db      69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10

Qy    1303 cctgaagta 1311
             : :: ::
Db       9 WWWWWWWWW 1


RESULT   7
AAF58262/c
ID   AAF58262 standard; DNA; 936 BP.
XX
AC   AAF58262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2007.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;


  Query Match            12.4%;  Score 219;  DB 22;  Length 936;
  Best Local Similarity   0.8%;  Pred. No. 1.7e-28;
  Matches    6;  Conservative  569;  Mismatches  214;  Indels    0;  Gaps    0;

Qy     523 gaaattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtgga 582
           | ::::: :  ::: : :::: ::  :  : :: :::::::: ::::::  :::  :  :
Db     789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730

Qy     583 gaaacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagctt 642
            :::  ::::: ::: ::::: ::::: ::::: :::::|:  :::::::::: ::  ::
Db     729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWW 670

Qy     643 gttaatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttac 702
            ::::: ::::: ::::: ::  : :: ::::: : :   ::  : ::   :: :::::
Db     669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610

Qy     703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
           :: :::::: ::::::: :::: :::::::::::::::::::::::: :: :::: ::::
Db     609 WWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550

Qy     763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
           : ::::    :: :::: ::::::: ::::: :  :: :: ::::  :  ::::: ::
Db     549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWW 490

Qy     823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
            ::: :  : ::  :: ::: : :::: :: :: ::: :  :::::: : :::::::::
Db     489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430

Qy     883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
           : : :: ::::: ::: :::::::: ::::: ::   ||:  :   :: :::: :  :  :
Db     429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWW 370

Qy     943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
           : ::::::::::::::::::::: :  :::  : :::::: :: :: :::::::: ::: :
Db     369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310

Qy    1003 tacaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatat 1062
           :: :::::: ::: :  ::::  : ::::::::  :: :::: :: ::::   ::::::::
Db     309 WWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250

Qy    1063 tttggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggt 1122
           :::  :::::::::::: :: : ::: ::::::  ::::: :::: :: ::::|: :  :
Db     249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWW 190

Qy    1123 ggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaa 1182
             ::::: :   :::::::: :::::::: :::: ::  ::::: :::: :: :::::
Db     189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130

Qy    1183 aatcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaa 1242
           ::: ::::::::::::    : ::  :::: | ::::   :::::::::: ::: :: ::
Db     129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWW 70

Qy    1243 ttaccaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaat 1302
           :::  :: :  :: :::  :: :: :::::::: ::: ::::::::  ::: ::::::::
Db      69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10

Qy    1303 cctgaagta 1311
             : :: ::
Db       9 WWWWWWWWW 1


RESULT   8
AAF58255/c
ID   AAF58255 standard; DNA; 938 BP.
XX
AC   AAF58255;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1876.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
```

```
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;


  Query Match             12.4%;  Score 219;  DB 22;  Length 938;
  Best Local Similarity   0.8%;  Pred. No. 1.7e-28;
  Matches    6;  Conservative  569;  Mismatches  214;  Indels    0;  Gaps    0;

Qy     523 gaaattagacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtgga 582
           | ::::: :  ::: : :::: ::  :  : :: :::::::: :::::  :::  :  :
Db     789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730

Qy     583 gaaacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagctt 642
            :::  ::::: ::: ::::: ::::: ::::: :::::|:  :::::::::: ::  ::
Db     729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWW 670

Qy     643 gttaatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttac 702
           :::::: ::::: ::::: ::  : :: ::::: : :   ::  : ::   :: :::::
Db     669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610

Qy     703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
           :: :::::: :::::::: :::: :::::::::::::::::::::::: :: :::: ::::
Db     609 WWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550

Qy     763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
           : ::::    :: :::: ::::::: ::::: :  :: :: ::::  :  :::::: ::
Db     549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWW 490

Qy     823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
           ::: :  : ::  :: ::: : :::: :: :: ::: :  :::::: : :::::::::
Db     489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430

Qy     883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
           : : :: ::::: ::: :::::::: ::::: ::     ||:  :   :: :::: :  :  :
Db     429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWW 370

Qy     943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
           : :::::::::::::::::::: :  :::  : :::::: :: :: :::::::: ::: :
Db     369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310

Qy    1003 tacaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatat 1062
           :: :::::: ::: :  ::::  : :::::::  :: :::: :: ::::   ::::::::
Db     309 WWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250

Qy    1063 tttggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggt 1122
           :::  :::::::::::: :: : ::: ::::::  ::::: :::: :: ::::|: :  :
Db     249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWW 190

Qy    1123 ggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaa 1182
           :::::  :    :::::::: :::::::: :::: ::  ::::: :::: :: :::::
Db     189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130

Qy    1183 aatcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaa 1242
           ::: ::::::::::::    : ::  :::: | ::::   :::::::::: ::: :: ::
Db     129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWW 70

Qy    1243 ttaccaactggatcattcgatgtagatatttttcttacaaatatttggaaacatttaaat 1302
           :::  :: :  :: :::  :: :: :::::::: ::: ::::::::  ::: ::::::::
Db      69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10

Qy    1303 cctgaagta 1311
           : :: ::
Db       9 WWWWWWWWW 1
```

```
RESULT   9
AAF58252
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;


  Query Match             12.3%;  Score 217.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 3.4e-28;
  Matches    3;  Conservative  569;  Mismatches  212;  Indels    0;  Gaps    0;

Qy       4 aattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccaga 63
           :::::::::::::: :::::::::::::::: :::  :: :: :::: :  :::::  : :
Db       3 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 62

Qy      64 accttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaat 123
           :  :: :::  :::::::::::::::::::: :::   :::::::    ::::::::::
Db      63 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWW 122

Qy     124 aaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaat 183
           :::: : :: :::   ::::: :: : : ::::: : :: ::::::  :::::::: ::::
Db     123 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 182

Qy     184 tttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataat 243
           ::: :: ::::: |:: :::::    : ::   : ::: : :: ::::: ::::::::::
Db     183 WWWWWWWWWWWWWWgWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 242

Qy     244 aaaaatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaa 303
           :::::::::::::::::: ::: ::::: :  :: : :::::::: :  :  ::::::::
Db     243 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWcWWWWWWW 302

Qy     304 aggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagca 363
```

```
              :  ::: ::  :::::::  :  ::: :::  :: ::  :: : : ::  : ::: :  :
Db       303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362

Qy       364 aaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatt 423
              ::: :::: ::: ::: :::      ::: :|    |::::::::::  ::::  :    :::
Db       363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 422

Qy       424 aatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaa 483
              ::::: :::::: :: ::: :: :  :::::::  :: : ::: :::  :::::: ::::
Db       423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482

Qy       484 cacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaa 543
               :     ::: : ::: :::  :::  ::: ::  ::::: :::::: :  ::: : ::::
Db       483 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542

Qy       544 gtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatata 603
               ::  :   : :: :::::::: ::::::  :::   :   : :::   ::::   ::: :::::
Db       543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 602

Qy       604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgatttgatttagat 663
               ::::: ::::: ::::: :  ::::::::: ::   :: ::::: ::::: ::::: ::
Db       603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662

Qy       664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
                : :: ::::: : :    ::   : ::    :: ::::: :: :::::: ::::::::::
Db       663 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722

Qy       724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
              : ::::::::::::::::::::: :: :::: ::::: ::::     :: :::: :::
Db       723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782

Qy       784 attt 787
              ::::
Db       783 wwww 786
```

RESULT 10
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

```
  Query Match             12.3%;  Score 217.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 3.4e-28;
  Matches    3;  Conservative  569;  Mismatches  212;  Indels    0;  Gaps    0;

Qy         4 aattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccaga 63
              ::::::::::::: :::::::::::::::: :::    :: :: :::: :   ::::: : :
Db         3 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 62

Qy        64 accttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaat 123
              :  :: :::   ::::::::::::::::::: :::.   :::::::     :::::::::::
Db        63 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwww 122

Qy       124 aaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaatatagtaat 183
              :::: : :: :::    ::::: :: : : ::::: : :: ::::::  ::::::: :.:::
Db       123 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 182

Qy       184 tttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataat 243
              ::: :: ::::: |:: :::::     : ::     : ::: : :: ::::: ::::::::::
Db       183 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 242

Qy       244 aaaaatataaataaaaatgatagaaaatcaccaagacaaatttagaggagtataaaaaa 303
              :::::::::::::::: ::: :::::  :   :: : ::::::::  :   :  ::::::::
Db       243 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwww 302

Qy       304 aggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagca 363
              :  ::: ::  :::::::  :  ::: :::  :: ::  :: : : ::  : ::: :  :
Db       303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362

Qy       364 aaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatt 423
              ::: :::: ::: ::: :::      ::: :|    |::::::::::  ::::  :    :::
Db       363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 422

Qy       424 aatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaa 483
              ::::: :::::: :: ::: :: :  :::::::  :: : ::: :.::  :::::: ::::
Db       423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482

Qy       484 cacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaa 543
               :     ::: : ::: :::  :::  ::: ::  ::::: :::::: :  ::: : ::::
Db       483 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542

Qy       544 gtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatata 603
               ::  :   : :: :::::::: ::::::  :::   :   : :::   ::::   ::: :::::
Db       543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 602

Qy       604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgatttgatttagat 663
               ::::: ::::: ::::: :  ::::::::: ::   :: ::::: ::::: ::::: ::
Db       603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662

Qy       664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
                : :: ::::: : :    ::   : ::    :: ::::: :: :::::: ::::::::::
Db       663 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722

Qy       724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
              : ::::::::::::::::::::: :: :::: ::::: ::::     :: :::: :::
Db       723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782

Qy       784 attt 787
              ::::
Db       783 wwww 786
```

RESULT 11
AAF58257
ID AAF58257 standard; DNA; 936 BP.

```
XX
AC   AAF58257;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

  Query Match             12.3%;  Score 217.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 3.4e-28;
  Matches    3;  Conservative  569;  Mismatches  212;  Indels    0;  Gaps    0;

Qy          4 aattttaaaatatcaatattttaattatagtatccatcttgtattctgcaaattccaga 63
              ::::::::::::: :::::::::::::: :::   :: :: :::: :  :::::  : :
Db          3 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 62

Qy         64 accttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaat 123
              :  :: :::  :::::::::::::::::: :::    :::::::    :::::::::::
Db         63 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwww 122

Qy        124 aaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaat 183
              :::: : :: :::   ::::: :: : : ::::: : :: ::::::  ::::::: ::::
Db        123 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 182

Qy        184 tttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataat 243
              ::: :: ::::: |:: :::::    : ::   : ::: : :: ::::: :::::::::
Db        183 wwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 242

Qy        244 aaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaa 303
              ::::::::::::::::: ::: ::::: :   :: : :::::::: :   :  ::::::::
Db        243 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwww 302

Qy        304 aggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagca 363
              :  ::: ::  :::::::  :  ::: :::  :: ::  :: : : ::  : ::: :  :
Db        303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362

Qy        364 aaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatt 423
              ::: :::: ::: ::: :::    :::  :|    |:::::::::::  ::::  :    :::
Db        363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwww 422

Qy        424 aatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaa 483
              ::::: :::::: :: ::: :: :  :::::::  :: : ::: :::  :::::: ::::
Db        423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482

Qy        484 cacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaa 543
               :      ::: : ::: :::  :::  ::: ::  ::::: :::::: :   ::: : ::::
Db        483 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542

Qy        544 gtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatata 603
              ::  :   : :: ::::::: ::::::  :::   :    : :::  ::::  ::: :::::
Db        543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 602

Qy        604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagat 663
              ::::: ::::: ::::: :  :::::::::: ::   :: ::::: ::::: ::::: ::
Db        603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662

Qy        664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
               : :: ::::: : :    ::   :  ::    :: ::::: :: :::::: ::::::::::
Db        663 wwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722

Qy        724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
              : ::::::::::::::::::::: :: :::: ::::: ::::    :: :::: :::
Db        723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782

Qy        784 attt 787
              ::::
Db        783 wwww 786


RESULT  12
AAF58259
ID   AAF58259 standard; DNA; 936 BP.
XX
AC   AAF58259;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2004.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
```

```
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;


  Query Match             12.3%;  Score 217.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 3.4e-28;
  Matches    3;  Conservative  569;  Mismatches  212;  Indels    0;  Gaps    0;

Qy      4 aattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccaga 63
          :::::::::::: ::::::::::::::: ::: :: :: :::: :  ::::: : :
Db      3 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 62

Qy     64 accttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaat 123
          :  :: :::  :::::::::::::::::::::: :::   :::::::    :::::::::::
Db     63 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 122

Qy    124 aaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaat 183
          :::: : :: :::   ::::: :: : : ::::: : :: :::::::  ::::::: ::::
Db    123 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 182

Qy    184 tttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataat 243
          ::: :: ::::: |:: :::::    : ::   : ::: : :: ::::: ::::::::::
Db    183 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 242

Qy    244 aaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaa 303
          ::::::::::::::::: ::: ::::: :  :: : :::::::: :  :  ::::::::
Db    243 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwww 302

Qy    304 aggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagca 363
          :  ::: ::  :::::::  :  ::: ::: :: ::  :: : : ::  : ::: :  :
Db    303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362

Qy    364 aaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatt 423
          ::: :::: ::: ::: :::   ::: :|   |:::::::::::  ::::  :  :::
Db    363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 422

Qy    424 aatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaa 483
          ::::: :::::: :: ::: :: :  :::::::  :: : ::: :::  :::::: ::::
Db    423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482

Qy    484 cacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaa 543
           :    ::: : ::: :::  :::  ::: ::  ::::: :::::: :  ::: : ::::
Db    483 wwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542

Qy    544 gtacgtccagatgtaattattctttttatccttaggtggagaaacctatatgatagatata 603
           ::  :  : :: :::::::: ::::::  :::  :  : :::  ::::  ::: :::::
Db    543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwww 602

Qy    604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagat 663
          :::::: ::::: ::::: :  :::::::::: ::  :: ::::: ::::: ::::: ::
Db    603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662

Qy    664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
            : :: ::::: : :   ::  : ::   :: ::::: :: :::::: ::::::::::
Db    663 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722

Qy    724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
          : :::::::::::::::::::::: :: :::: ::::: ::::    :: :::: :::
Db    723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782

Qy    784 attt 787
          ::::
Db    783 wwww 786


RESULT  13
AAF58262
ID   AAF58262 standard; DNA; 936 BP.
XX
AC   AAF58262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2007.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;


  Query Match             12.3%;  Score 217.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 3.4e-28;
  Matches    3;  Conservative  569;  Mismatches  212;  Indels    0;  Gaps    0;

Qy      4 aattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccaga 63
          :::::::::::: ::::::::::::::: ::: :: :: :::: :  ::::: : :
Db      3 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 62

Qy     64 accttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaat 123
          :  :: :::  :::::::::::::::::::::: :::   :::::::    :::::::::::
Db     63 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 122

Qy    124 aaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaat 183
          :::: : :: :::   ::::: :: : : ::::: : :: :::::::  ::::::: ::::
Db    123 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 182

Qy    184 tttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataat 243
          ::: :: ::::: |:: :::::    : ::   : ::: : :: ::::: ::::::::::
Db    183 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 242

Qy    244 aaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaa 303
          ::::::::::::::::: ::: ::::: :  :: : :::::::: :  :  ::::::::
Db    243 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwww 302

Qy    304 aggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagca 363
          :  ::: ::  :::::::  :  ::: ::: :: ::  :: : : ::  : ::: :  :
Db    303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362

Qy    364 aaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatt 423
          ::: :::: ::: ::: :::   ::: :|   |:::::::::::  ::::  :  :::
Db    363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 422

Qy    424 aatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaa 483
          ::::: :::::: :: ::: :: :  :::::::  :: : ::: :::  :::::: ::::
Db    423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482
```

```
Qy    484 cacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaa 543
          :    ::: : ::: :::  :::  ::: ::  ::::: :::::: :  ::: : ::::
Db    483 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542

Qy    544 gtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatata 603
           ::  :  : :: :::::::: ::::::  :::  :  : :::  ::::  ::: :::::
Db    543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwww 602

Qy    604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagat 663
          :::::  ::::: ::::: :  :::::::::: ::  :: ::::: ::::: ::::: ::
Db    603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662

Qy    664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
            : :: ::::: : :   ::  : ::   :: ::::: :: :::::: ::::::::::::
Db    663 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722

Qy    724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
          : ::::::::::::::::::::::: :: :::: ::::: ::::    :: :::: :::
Db    723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782

Qy    784 attt 787
          ::::
Db    783 wwww 786
```

RESULT 14
AAF58255
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 12.3%; Score 217.2; DB 22; Length 938;
Best Local Similarity 0.4%; Pred. No. 3.4e-28;
Matches 3; Conservative 569; Mismatches 212; Indels 0; Gaps 0;

```
Qy      4 aattttaaaatatcaatatttttaattatagtatccatcttgtattctgcaaattccaga 63
          :::::::::::: :::::::::::::::: :::  :: :: :::: :  :::::  : :
Db      3 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 62

Qy     64 accttgaaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaat 123
          :  :: :::  :::::::::::::::::::: :::   :::::::    ::::::::::
Db     63 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 122

Qy    124 aaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaat 183
          :::: : :: :::  ::::: :: : : ::::: : :: ::::::  ::::::: ::::
Db    123 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 182

Qy    184 tttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataat 243
          ::: :: ::::: |:: :::::    : ::   : ::: : :: ::::: ::::::::::
Db    183 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 242

Qy    244 aaaaatataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaa 303
          :::::::::::::::::: ::: ::::: :  :: : :::::::: :  :  ::::::::
Db    243 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwww 302

Qy    304 aggaaacaaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagca 363
          :  ::: ::  :::::::  :  ::: :::  :: ::  :: : : ::  : ::: :  :
Db    303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362

Qy    364 aaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcatt 423
          ::: :::: ::: ::: :::   :::  :|   |:::::::::::  ::::  :   :::
Db    363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwww 422

Qy    424 aatatgttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaa 483
          ::::: :::::: :: ::: :: :  :::::::  :: : ::: :::  :::::: ::::
Db    423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482

Qy    484 cacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaa 543
          :    ::: : ::: :::  :::  ::: ::  ::::: :::::: :  ::: : ::::
Db    483 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542

Qy    544 gtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgatagatata 603
           ::  :  : :: :::::::: ::::::  :::  :  : :::  ::::  ::: :::::
Db    543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwww 602

Qy    604 gaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagat 663
          :::::  ::::: ::::: :  :::::::::: ::  :: ::::: ::::: ::::: ::
Db    603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662

Qy    664 ggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttt 723
            : :: ::::: : :   ::  : ::   :: ::::: :: :::::: ::::::::::::
Db    663 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722

Qy    724 tcaaattattatattaaattaattaacttgttaagaaaaactattccggaagaaaagtta 783
          : ::::::::::::::::::::::: :: :::: ::::: ::::    :: :::: :::
Db    723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782

Qy    784 attt 787
          ::::
Db    783 wwww 786
```

RESULT 15
AAC89668
ID AAC89668 standard; DNA; 1137 BP.
XX
AC AAC89668;
XX
DT 13-MAR-2001 (first entry)
XX
DE P. falciparum chitinase coding sequence SEQ ID NO: 1.
XX
KW Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KW veterinary disease; arthropod pest; ds.
XX

OS Plasmodium falciparum.
XX
PN WO200073488-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14536.
XX
PR 28-MAY-1999; 99US-0136508.
PR 03-FEB-2000; 2000US-0180051.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Vinetz JM;
XX
DR WPI; 2001-061553/07.
DR P-PSDB; AAB49985:
XX
PT New nucleic acid encoding a Plasmodium species chitinase is useful for
PT preventing transmission of malaria by mosquito feeding on subject that
PT may harbor Plasmodium species organisms -
XX
PS Claim 4; Page 123; 137pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC organisms are the cause of malaria in humans. The sequences are useful
CC in the prevention and treatment of malaria, fungal diseases, parasitic
CC diseases and veterinary diseases, in preventing the transmission of
CC malaria and in the control of arthropod pests in agriculture.
XX
SQ Sequence 1137 BP; 321 A; 263 C; 257 G; 296 T; 0 other;

Query Match 5.9%; Score 104.2; DB 22; Length 1137;
Best Local Similarity 48.0%; Pred. No. 2.9e-09;
Matches 468; Conservative 0; Mismatches 483; Indels 24; Gaps 5;

```
Qy    266 gaaaatcaccaagacaaatttagaggagtataaaaaaaggaaacaaggtattatagcag 325
Db    110 gcaaaaacccgcgtgaaattatcaaaacgtttaaagaatccggtaaaggtatcatccagg 169

Qy    326 gatactatggttcatggaacagtcaaggtgatagagcaaaacatatgattgattcaaacc 385
Db    170 gttactatccgtcctgggtgagctataatcaca---------acctgaaggacctgaacc 220

Qy    386 caatggtgtcaattttatatattgcatttgctcgcattaatatgttatatgatgtatcta 445
Db    221 cgaacctgaacgtggttcacatgagctttgctaaaatggatttaagctacgactccattg 280

Qy    446 gaccatttaatggaagacaaagattcctattaagaaaacacggcttagaatatgaaacct 505
Db    281 aatccatcgtaggtagcccgctgctgtttaaaagcctgattggcctggagtatattggtc 340

Qy    506 atggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagatgtaattattc 565
Db    341 tgaatgagtacttcaatgatgccatgaatctgcgcaaagctcgtccggacattattatgc 400

Qy    566 ttttatccttaggtggagaaacct---atatgatagatatagaaaaagaaattgattatg 622
Db    401 tgctgagcctgggcggtgagacctaccatccgtcctccttcgattctgcactcaacgcgg 460

Qy    623 tggataaaatattgaagcttgttaatgattttgatttagatggtgtagatattgactggg 682
Db    461 ttgaaaaaatcgcaaatctggtggatgaactgggcttcgatggtattgatgtagattacg 520

Qy    683 aaccacatgggaagttttacaacttaaatgaattaaatttttcaaattattatattaaat 742
Db    521 aaccgaacggcagctttgatggtctgaatgataaagaaaaagcggacttctttgtacaat 580

Qy    743 taattaacttgttaagaaaaactattccggaagaaaagttaatttcaatttctggttcat 802
Db    581 acgtgaccaaactgcgcgaatacatgtgtgatgataaactgatcagcatcagccagtcct 640

Qy    803 caaatgctgcattatcatgcgtttcaggagttgcatctt------tctgtaaagatgaag 856
Db    641 ctaatggcgctctgagctgcatcggtttcaacgacccgaaaaaaatctgtatggatgacg 700

Qy    857 aatctccatataacactaaatttttgtctgaacaaatagaaacaaataaagaattacata 916
Db    701 aagctccgtataacagcaaatattt---caacaaaccggacgttaagaaagaactgttac 757

Qy    917 gggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaaggaga 976
Db    758 gcgcagcccagatggcatctgcgggtggtgccatctacctgatgaacaacctgaaagata 817

Qy    977 aaatagatcttgtatttattcaaacatacaatt---tagaaactacaaatccagatataa 1033
Db    818 tgattgacatggtgtttgtgcagacgttcaactacacgaactctaccgattctaccgtta 877

Qy   1034 tggtagatatgtacttatcccatttatattttggtttaaaatataacatcacaatcatat 1093
Db    878 tgaaagaattatacgactcctacgcatactatggtaaaaagtacgattacgtgatcatca 937

Qy   1094 taggtttttcattagaacataacagaggtggatttagtcccgaaaataaagaattattag 1153
Db    938 tgggcttcaccctgatgttcccgtccacgccgttcaacccgaacgataaaatgctggtaa 997

Qy   1154 aattggtaggaaaaacaatacatgataaaaatcaaaataataatagggcagatggtatag 1213
Db    998 aatctatcggcgatttcgtaaaaactgaaaacaaactgaataaacgcgcagatggtttcg 1057

Qy   1214 ggatatggcatttat 1228
Db   1058 gcctgtggtctctgt 1072
```

Search completed: March 21, 2002, 15:43:51
Job time: 4462 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:28:34 ; Search time 94.72 Seconds
(without alignments)
4217.770 Million cell updates/sec

Title: US-09-579-383-2
Perfect score: 1764
Sequence: 1 atgaattttaaaatatcaat.....tggatttaccattacaataa 1764

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	62.2	3.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	2	60.6	3.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	3	60	3.4	834	4	US-08-998-416-305	Sequence 305, App
c	4	59.2	3.4	837	4	US-08-998-416-288	Sequence 288, App
c	5	58	3.3	615	4	US-08-998-416-186	Sequence 186, App
c	6	58	3.3	665	2	US-08-883-795A-36	Sequence 36, Appl
c	7	57	3.2	660	1	US-07-991-867B-32	Sequence 32, Appl
c	8	57	3.2	660	1	US-08-107-755A-32	Sequence 32, Appl
c	9	57	3.2	660	2	US-08-544-332-32	Sequence 32, Appl
c	10	57	3.2	1511	1	US-07-991-867B-8	Sequence 8, Appli
c	11	57	3.2	1511	1	US-08-107-755A-8	Sequence 8, Appli
c	12	57	3.2	1511	2	US-08-544-332-8	Sequence 8, Appli
c	13	57	3.2	4810	3	US-08-852-629-11	Sequence 11, Appl
c	14	57	3.2	4838	3	US-08-852-629-15	Sequence 15, Appl
	15	54.4	3.1	5852	1	US-07-867-106-2	Sequence 2, Appli
	16	53.6	3.0	3095	6	5231168-1	Patent No. 5231168
c	17	53.4	3.0	636	4	US-08-998-416-1137	Sequence 1137, Ap
c	18	53	3.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
	19	52.4	3.0	665	2	US-08-883-795A-36	Sequence 36, Appl
	20	52	2.9	8920	2	US-08-446-855A-1	Sequence 1, Appli
	21	52	2.9	8920	4	US-09-150-741-1	Sequence 1, Appli
c	22	51.2	2.9	5852	1	US-07-867-106-2	Sequence 2, Appli
	23	49.8	2.8	240	1	US-08-628-417-6	Sequence 6, Appli
c	24	49	2.8	1689	1	US-07-991-867B-41	Sequence 41, Appl
c	25	49	2.8	1689	2	US-08-544-332-41	Sequence 41, Appl
	26	49	2.8	6243	2	US-09-056-075-1	Sequence 1, Appli
c	27	49	2.8	8457	1	US-07-991-867B-1	Sequence 1, Appli
c	28	49	2.8	8457	2	US-08-544-332-1	Sequence 1, Appli
	29	48.6	2.8	1066	1	US-08-157-101A-4	Sequence 4, Appli
c	30	48.2	2.7	724	4	US-09-007-119-8	Sequence 8, Appli
	31	47.8	2.7	828	4	US-08-998-416-538	Sequence 538, App
	32	47.6	2.7	701	4	US-08-998-416-701	Sequence 701, App
	33	47.4	2.7	854	4	US-08-998-416-534	Sequence 534, App
	34	47.4	2.7	2230	3	US-08-844-188-34	Sequence 34, Appl
	35	47.4	2.7	2447	2	US-09-014-969-14	Sequence 14, Appl
c	36	47.4	2.7	9048	3	US-08-973-273-4	Sequence 4, Appli
c	37	47.2	2.7	731	1	US-08-451-405A-2	Sequence 2, Appli
c	38	47.2	2.7	1683	4	US-09-009-443-6	Sequence 6, Appli
c	39	46.8	2.7	3701	4	US-08-845-258-10	Sequence 10, Appl
c	40	46.8	2.7	3701	4	US-08-990-571-10	Sequence 10, Appl
c	41	46.8	2.7	3701	4	US-08-723-142A-10	Sequence 10, Appl
c	42	46.4	2.6	625	4	US-08-998-416-919	Sequence 919, App
c	43	46.4	2.6	713	4	US-08-998-416-956	Sequence 956, App
	44	46.4	2.6	1431	4	US-09-316-083-2	Sequence 2, Appli
c	45	46	2.6	1736	3	US-09-182-816-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 3.5%; Score 62.2; DB 2; Length 19124;

```
  Best Local Similarity   43.0%;  Pred. No. 0.00033;
  Matches  304;  Conservative    0;  Mismatches  403;  Indels    0;  Gaps    0;

Qy     629 aaatattgaagcttgttaatgattttgatttagatggtgtagatattgactgggaaccac 688
           | |||| || |    |  |||  ||| || || ||  | || ||||        |
Db   15360 ATATATGGATGTACTTGTATGTGTTTTATATATATATTTTATATATATGTATTTATATTA 15419

Qy     689 atgggaagttttacaacttaaatgaattaaatttttcaaattattatattaaattaatta 748
           |     |  | || ||   || | | | ||||     |||  |  ||   | || ||  |
Db   15420 AAAAAGAAATATAAAAACAAATTTATTAAAATGAAAAAAAGAAAAATGAAATATAAAAAA 15479

Qy     749 acttgttaagaaaaactattccggaagaaaagttaatttcaatttctggttcatcaaatg 808
           |  | |    ||| |  |      || ||||   |     ||  | |  |  |  | | |
Db   15480 AAATTTATTAAAATAAAAAAAAAAAAAAAAAAAAAAGGAGAAAAATTTTTTAAAAAATAAT 15539

Qy     809 ctgcattatcatgcgtttcaggagttgcatctttctgtaaagatgaagaatctccatata 868
               ||||| ||    |  |    |||      |    || ||  |||| | || | | |
Db   15540 AAAAATTATAATAAAATATAAATTTTGATAGAATAAAAAATGAAAAAGATTATCAAAAAA 15599

Qy     869 acactaaatttttgtctgaacaaatagaaacaaatgaagaattacatagggcagcagcga 928
           | | ||||      | | |   || | ||| |     | ||  | |      |  |  ||
Db   15600 AAATTAAAAAAAAATTTTATATAAAAAAAAAATGATTATAAAAAAAATAAAAACAAAAGA 15659

Qy     929 tgttatcagcaggaacttttattaatatttttaatacagcaaaggagaaaatagatcttg 988
           |  |  |  |   | |   |  || |   |  |||    |||    |||| | |
Db   15660 AGAAAAAAAAAAACATTAAAAAAAAAAAAAATATATATCATAAAAACAAAAAAAAAAGAAA 15719

Qy     989 tatttattcaaacatacaatttagaaactacaaatccagatataatggtagatatgtact 1048
           |  |||   || ||| || | |    | |   ||||  | | | | |  | |   ||
Db   15720 AAAATATATTAAAATAAAAATATATATCATAAAATAAAAAAAAATTAAAAAAATGTTAAA 15779

Qy    1049 tatcccatttatattttggtttaaaatataacatcacaatcatattaggtttttcattag 1108
           |    || ||||  |   | ||||  |    ||   ||| | |  |  |  |  || |
Db   15780 AAAAAAATATATACATAAAATAAAAAAAATTTATTTAAATAAAAAAAAATAATAAATAAA 15839

Qy    1109 aacataacagaggtggatttagtcccgaaaataaagaattattagaattggtaggaaaaa 1168
           || ||   |       |   |       |||||||  || | | |  |      |  ||||
Db   15840 AAAATTTAATTAAATAAAAAAAAATAATAAATAAAAAAATTTAATTAAATAAAAAAAAT 15899

Qy    1169 caatacatgataaaaatcaaaataataataggggcagatggtatagggatatggcatttat 1228
           || | |   |||  |   ||| || ||||    |  |    ||   | |    ||  |
Db   15900 TAAAAAAATTTAATGAAATAAAAAAAAATAAAAAAATTTAATTAAATAAAAAAAATAAAA 15959

Qy    1229 ttatgaaagaacaattaccaactggatcattcgatgtagatatttttcttacaaatattt 1288
           | |    |      ||   ||  |  ||   |  || || |||| |  |    |     |
Db   15960 TAAAATTAATTACATGCACATATACATACATATATATATATATATACCCATAACTACATA 16019

Qy    1289 ggaaacatttaaatcctgaagtacaaactccaaaagaccttactata 1335
             | |  |  | || |  |  || | |     | |  | | |||| |
Db   16020 CAACATTTACACATACATATATATATATATATACCCATAACTACA 16066


RESULT   2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             3.4%;  Score 60.6;  DB 1;  Length 7218;
  Best Local Similarity   3.9%;  Pred. No. 0.00058;
  Matches   15;  Conservative  224;  Mismatches  148;  Indels    0;  Gaps    0;

Qy    1363 atagatgaatatgttccaggactcgttattccaaccatagggatatattacaaacacaat 1422
           ||||| |||| || | |::::   :  :    ::  : ::::: : :  : ::: : ::
Db    1451 ATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392

Qy    1423 gatgctatatggaaaactagatcttattcaattcatgcacctggtgtagacagatatgaa 1482
           :: :  : : :::::  :::    :   ::   : : :   :: : ::: ::: : :::
Db    1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332

Qy    1483 tgggacttggtcaaagtatgctatgaaaaaatatgcgatgggaaagcagcccattattat 1542
            ::::   ::  :::: : :  : ::::::: : : :: ::::::: ::   :  :  :
Db    1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272

Qy    1543 aacactgactataaagaaagctctattattatatggaaaggggaaccatatttaattaaa 1602
           :: :  ::  : ::::::::    :  :  : : ::::::::::::  : :   ::  :::
Db    1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212

Qy    1603 tggtggcaacaaggacctccggaaggtcaggcactagagtcatacacaaaactagatgca 1662
           :: :: :: :::::     ::::::  ::: :  ::::  : : : ::::  ::: : :
Db    1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152

Qy    1663 tccaaatgtccagggatagaagaatggaataaaaaatatccacataaaccactagaagta 1722
              ::: :   ::::: ::::::: :::: :::::: :   : : :::  :  ::::: :
Db    1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092

Qy    1723 gaggaacaatatgaacaagaagtggat 1749
           :::::: :: : ::: :::::: ::||
Db    1091 RRRRRRRRRRRRRRRRRRRRRRRRRAT 1065


RESULT   3
US-08-998-416-305
; Sequence 305, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
```

```
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  305:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  834 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1250RP
US-08-998-416-305


  Query Match             3.4%;  Score 60;  DB 4;  Length 834;
  Best Local Similarity   45.9%;  Pred. No. 0.0005;
  Matches  275;  Conservative    0;  Mismatches  321;  Indels    3;  Gaps    2;

Qy    478 agaaaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatc 537
          | || | |  |   |  ||| |  | |     || | |   |||| |||||   ||  |
Db    114 AAAATAAATAGAAAACCATAAGTTAATTGATTCATAAAGAAAAATGGAATTATTTGTGGC 173

Qy    538 agaaaagtacgtccagatgtaattattcttttatccttaggtggagaaacct--atatga 595
          |   | |  |     |  | | ||| || | |  |||     | | |  |  | |||
Db    174 ATCTTAATTTTTATTATTTAATTGATTATTATCTATTTAACATAAAACATTTTAAAATGT 233

Qy    596 tagatatagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattttg 655
          || | |   || ||||||||  ||||         ||    ||  | | | | | |||
Db    234 TATAAAATAAATAAGAAATTACTTATAGAATATTTATTAAATAGTATTTAATTTAATTTT 293

Qy    656 atttagatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatgaat 715
          | |   |      || || | |   ||  | ||    |  |||| | |   | |
Db    294 AATATTAAATATACCATTTTTATTAATAAATAGATTATTAAGTTTATTAATATTAAGTGA 353

Qy    716 taaatttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaag 775
          || ||  || || |||   | | | |||||| ||||  |    | |  || |    ||
Db    354 TATATAATTTAATTTATATAAATTATTTAATTTACTTCATTGATATATATAATTATTAAA 413

Qy    776 aaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttg 835
            |  |    | |  ||    || ||  | |   |.| |||      ||    | |
Db    414 TGTACCTTTCATAATATTTATTTTTATT-AGTCTAGTAATATTTCTATTTAATAGTCTAC 472

Qy    836 catctttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatag 895
          | | |   || | |     |   || ||||   ||||   | | |    |  ||||
Db    473 CCTTTAATTGGATATTACTACCTACTAAATATTTACCTAATAATATATTATTAAGAATAC 532

Qy    896 aaacaaataaagaattacatagggcagcagcgatgttatcagcaggaacttttattaata 955
            | |  |||  | |||        || |   |  |||     |     |||||||| ||
Db    533 TTAAATCTAATAATTTATTATCTAAAGTATATAAATTAATTAAATCCTTTTTTATTATTA 592

Qy    956 tttttaatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaa 1015
          |||       |  ||     ||| || || |  ||||  | ||    |  |||   ||
Db    593 TTTAATTATTATTAATTAGTAAATTATATTTATTATTTTATTAACATAATTTTTTGATAA 652

Qy   1016 ctacaaatccagatataatggtagatatgtacttatcccatttatattttggtttaaaa 1074
            | | ||||| ||  |||||||  |   || | ||  | ||||    ||   | | ||
Db    653 TAATATATCCATATTAAATGGTAATTTATTAATAATATCCTTTAATGATTTNATGATAA 711


RESULT   4
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  288:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1241RP
US-08-998-416-288


  Query Match             3.4%;  Score 59.2;  DB 4;  Length 837;
  Best Local Similarity   45.1%;  Pred. No. 0.00073;
  Matches  259;  Conservative    0;  Mismatches  313;  Indels    2;  Gaps    1;

Qy    492 agaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtcc 551
          | |||   |||    |  |   ||  |||||  ||||| |   || |  |||  |  |
```

```
Db    585 ATAATGATAAAATATTAATTAAATAATTAATTTAATTAAATAAATAAATAAATAAAATGA 526

Qy    552 agatgtaattattctttttatccttaggtggagaaacctatatgatagatatagaaaaaga 611
          |     | |||||   | ||    |  |  | ||    | || | || || ||  |    ||
Db    525 AATATAATTTATTAAATAATATAAAAATTAATAATAAGAAATTAAAGTTAAAATTAATTT 466

Qy    612 aattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgtaga 671
          | | || ||    ||||||   | ||   |  ||   ||    ||     |  |  | |
Db    465 TAATAATAATTCTTATAAAAAGATTAA--ATAATATAAATCAACATAATATTTATAAAAA 408

Qy    672 tattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaaatta 731
          ||   | |   |   | ||   || ||  |  | |||||| |||||||| |||  ||| |
Db    407 TAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAATTAAATCTTTATAATAA 348

Qy    732 ttatattaaattaattaacttgttaagaaaaactattccggaagaaaagttaatttcaat 791
          | | ||  | |   ||||      |  || || || |     |   |     | |  | |
Db    347 TAAAATATATTATTTTAATAAACAATTAATAAATAATATATTAATTATTGATAATCTATT 288

Qy    792 ttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttctgtaaaga 851
          |  |  || || |||      | ||  ||    |       |   | ||   |  |||
Db    287 TAATAATTTATTAAAGAAAATAATAATATCTAATAATATTTTAATAACTAATTTAAAATT 228

Qy    852 tgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaaatagaaatt 911
          ||||  |    | | |||  | | |  ||   | | |        |    |  ||
Db    227 TGAACATAGACTAAATAGTATTCATATTAAATATTATTTTTATAATTATATAAATATTAA 168

Qy    912 acatagggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaa 971
            ||   | |   |  | ||||          | | || | |||| ||||  |  ||
Db    167 TAATGATGAATTAAGTAAATTATATAATAATTATATAATAAGTATTAATAATCAAATAAT 108

Qy    972 ggagaaaatagatcttgtatttattcaaacatacaatttagaaactacaaatccagatat 1031
          |   | || ||  ||    || |     | | ||||||  |  || | ||    |||
Db    107 TAATTTATTAAATAATGATAATAATAGTTTAAATAATTTAATACTTAAATATTATAATAA 48

Qy   1032 aatggtagatatgtacttatcccatttatatttt 1065
          ||  ||  ||||  |  | |   | ||| ||| |
Db     47 AAAAGTTTATATTAATCTTTATAAATTAAATTAT 14
```

RESULT 5
US-08-998-416-186/c
; Sequence 186, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1074RP
US-08-998-416-186

Query Match 3.3%; Score 58; DB 4; Length 615;
Best Local Similarity 46.2%; Pred. No. 0.0012;
Matches 265; Conservative 0; Mismatches 305; Indels 4; Gaps 2;

```
Qy    492 agaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtcc 551
          | |||   |||     |  |   ||  |||||  |||||  |   ||  |  |||  |  |
Db    583 ATAATGATAAAATATTAATTAAATAATTAATTTAATTAAATAAATAAATAAATAAAATGA 524

Qy    552 agatgtaattattctttttatccttaggtggagaaacctatatgatagatatagaaaaaga 611
          | ||  | |||||    | ||     |  |  | ||     | || | || || |    ||
Db    523 ATAT-AATTTATTAAATAATATAAAAATTAATAATAAGAAATTAAAGTTAAAATTAATTT 465

Qy    612 aattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgtaga 671
          | | || ||    ||||||  || ||    | | ||| |   | || |   |  | | |
Db    464 TAATAATAATTCTTATAAAAAGTTAAATAATATAAATCAACATAATAT---TTATAAAAA 408

Qy    672 tattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaaatta 731
          ||   | |   |   | ||   || ||  |  | |||||| |||||||| |||  ||| |
Db    407 TAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAATTAAATCTTTATAATAA 348

Qy    732 ttatattaaattaattaacttgttaagaaaaactattccggaagaaaagttaatttcaat 791
          | | ||  | |   ||||      |  || || || |     |   |     | |  | |
Db    347 TAAAATATATTATTTTAATAAACAATTAATAAATAATATATTAATTATTGATAATCTATT 288

Qy    792 ttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttctgtaaaga 851
          |  |  || || |||      | ||  ||    |       |   | ||   |  |||
Db    287 TAATAATTTATTAAAGAAAATAATAATATCTAATAATATTTTAATAACTAATTTAAAATT 228

Qy    852 tgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaaatagaaatt 911
          ||||  |    | | |||  | | |  ||   | | |        |    |  ||
Db    227 TGAACATAGACTAAATAGTATTCATATTAAATATTATTTTTATAATTATATAAATATTAA 168

Qy    912 acatagggcagcagcgatgttatcagcaggaacttttattaatatttttaatacagcaaa 971
            ||   | |   |  | ||||          | | || | |||| ||||  |  ||
Db    167 TAATGATGAATTAAGTAAATTATATAATAATTATATAATAAGTATTAATAATCAAATAAT 108

Qy    972 ggagaaaatagatcttgtatttattcaaacatacaatttagaaactacaaatccagatat 1031
          |   | || ||  ||    || |     | | ||||||  |  || | ||    |||
Db    107 TAATTTATTAAATAATGATAATAATAGTTTAAATAATTTAATACTTAAATATTATAATAA 48

Qy   1032 aatggtagatatgtacttatcccatttatatttt 1065
          ||  ||  ||||  |  | |   | ||| ||| |
Db     47 AAAAGTTTATATTAATCTTTATAAATTAAATTAT 14
```

RESULT 6
US-08-883-795A-36/c
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator

```
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BERESKIN & PARR
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,795A
;      FILING DATE:  27-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  7841-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 364-7311
;      TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  665 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Rh 32
US-08-883-795A-36


  Query Match             3.3%;  Score 58;  DB 2;  Length 665;
  Best Local Similarity   45.8%;  Pred. No. 0.0012;
  Matches  241;  Conservative    0;  Mismatches  280;  Indels    5;  Gaps    1;

Qy     585 aacctatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagcttgt 644
           ||  |||   ||    | | |   |   ||  ||  | |   || || ||      || |
Db     561 AAAATATTGAATTATAAAATATGTAATTATAAATACTTTAATTATAAAATATGTAATTAT 502

Qy     645 taatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttacaa 704
            |||  |||   | || |   ||||  |||  |      |  | ||  | | || || ||
Db     501 AAATACTTTAATTATAAAATATGTAATTATAAATACTTTATAAAATATGTAATTATAAAA 442

Qy     705 cttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaac 764
             |  |   || ||  |||| |   ||  ||||  |||||   |  || | |  | |||
Db     441 TATGTAATTATAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAA 382

Qy     765 tattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgt 824
           |||      | |||  ||    | |     |   | || | |  |    | || ||
Db     381 TATGTAATTATAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAA 322

Qy     825 ttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtc 884
           |     | |   | | || |     | | |     |    ||| |    | |||| |
Db     321 TATGTAATTATAAACATTTTAATTATAAAATATTTAATTATAAACATTTTAATTATAAAA 262

Qy     885 tgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaac 944
           |    || || ||| |  | ||  || | |||    |  |   || || | |  |  ||
Db     261 TATTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAAA 202

Qy     945 t-----tttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaa 999
           |      |||| |||||||| | || |  | |     |  ||| || || || ||||  ||
Db     201 TATTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAAA 142

Qy    1000 acatacaatttagaaactacaaatccagatataatggtagatatgtacttatcccattta 1059
             | | || | | |  || | | |     | |     |    |||           |
Db     141 TATTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAAA 82

Qy    1060 tattttggtttaaaatataacatcacaatcatattaggtttttcat 1105
           |||||   | ||||   |   || | ||  || |||  | |   ||
Db      81 TATTTAATTATAAATATTTTAATTATAAATATTTTAATTATAAAAT 36


RESULT   7
US-07-991-867B-32/c
; Sequence 32, Application US/07991867B
; Patent No. 5476781
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5476781el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,867B
;      FILING DATE:  12-DEC-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-991-867B-32


  Query Match             3.2%;  Score 57;  DB 1;  Length 660;
  Best Local Similarity   45.9%;  Pred. No. 0.0019;
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy     706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
           || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |     |
Db     629 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 570

Qy     766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                     || |  || | | ||| |    |||  || | |  ||     |   ||
Db     569 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 510

Qy     826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
             |  | ||   | ||     ||     ||   | |   | ||    |    | ||
Db     509 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 450
```

```
Qy    886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
Db    449 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 390

Qy    946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
Db    389 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 330

Qy   1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
Db    329 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 270

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
Db    269 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 210

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
Db    209 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 150

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
Db    149 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 91

Qy   1245 accaactggatcattcgatgtagatatttttcttaca 1281
Db     90 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 54


RESULT   8
US-08-107-755A-32/c
; Sequence 32, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-107-755A-32


  Query Match             3.2%;  Score 57;  DB 1;  Length 660;
  Best Local Similarity   45.9%;  Pred. No. 0.0019;
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy    706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
Db    629 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 570

Qy    766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
Db    569 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 510

Qy    826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
Db    509 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 450

Qy    886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
Db    449 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 390

Qy    946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
Db    389 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 330

Qy   1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
Db    329 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 270

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
Db    269 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 210

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
Db    209 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 150

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
Db    149 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 91

Qy   1245 accaactggatcattcgatgtagatatttttcttaca 1281
Db     90 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 54


RESULT   9
US-08-544-332-32/c
; Sequence 32, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-544-332-32


  Query Match             3.2%;  Score 57;  DB 2;  Length 660;
  Best Local Similarity   45.9%;  Pred. No. 0.0019;
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy     706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
           || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |     |
Db     629 TTTAAATAATTATAAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 570

Qy     766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                   || |  || | | ||| |    |||  || | |  ||     |   ||
Db     569 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 510

Qy     826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
             | | ||   | ||     ||     ||   | |   | ||    |    | ||
Db     509 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 450

Qy     886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
             |   |||  || |   |||  ||||   | |  | ||    | || | ||     | |
Db     449 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 390

Qy     946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
             | ||  ||  |||| || |  |      ||| || | || | ||| |      |   |
Db     389 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 330

Qy    1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
           ||||||| ||| |  | |    |   || |   | ||   |  |  | | |||| | ||
Db     329 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 270

Qy    1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
           |    || |    || ||   |    | |||  ||| | || | |    ||   |
Db     269 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 210

Qy    1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
             || | |    || ||| | | || | | | ||  | | | |     || |  || | |
Db     209 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 150

Qy    1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
           ||| ||| ||| ||      ||  |       ||    |||| |||| | | |  |||
Db     149 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTTAATATATTAATATT 91

Qy    1245 accaactggatcattcgatgtagatatttttcttaca 1281
           |  |||  ||| | |  | |  ||||||| ||| |
Db      90 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 54


RESULT  10
US-07-991-867B-8/c
; Sequence 8, Application US/07991867B
; Patent No. 5476781
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5476781el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,867B
;      FILING DATE:  12-DEC-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-07-991-867B-8


  Query Match             3.2%;  Score 57;  DB 1;  Length 1511;
  Best Local Similarity   45.9%;  Pred. No. 0.0023;
```

```
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy     706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
           || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |     |
Db    1480 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 1421

Qy     766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                   || |  || | | ||| |    |||  || | |  ||     |   ||
Db    1420 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 1361

Qy     826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
             |  | ||   | ||     ||     ||   | |   | ||    |    | ||
Db    1360 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 1301

Qy     886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
             |   |||  || |   ||| ||||   | |  | ||    | || | ||     | |
Db    1300 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 1241

Qy     946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
             | ||  ||  |||| || |  |      ||| || | || | ||| |      |   |
Db    1240 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 1181

Qy    1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
           ||||||| ||| |  | |    |   || |   | ||   |  |  | | |||| | ||
Db    1180 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 1121

Qy    1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
           |    || |    || ||   |    | |||  ||| | || | |    ||   |
Db    1120 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 1061

Qy    1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
             || | |    || ||| | | || | | ||  | | |     || |  || |   |
Db    1060 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 1001

Qy    1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
           ||| ||| ||| ||      ||  |       ||    |||| |||  | |  |  |||
Db    1000 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 942

Qy    1245 accaactggatcattcgatgtagatatttttcttaca 1281
           |  |||   |||  |   | |   |||||||| ||| |
Db     941 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 905


RESULT  11
US-08-107-755A-8/c
; Sequence 8, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-107-755A-8


  Query Match             3.2%;  Score 57;  DB 1;  Length 1511;
  Best Local Similarity   45.9%;  Pred. No. 0.0023;
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy     706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
           || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |     |
Db    1480 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 1421

Qy     766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                   || |  || | | ||| |    |||  || | |  ||     |   ||
Db    1420 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 1361

Qy     826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
             |  | ||   | ||     ||     ||   | |   | ||    |    | ||
Db    1360 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 1301

Qy     886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
             |   |||  || |   ||| ||||   | |  | ||    | || | ||     | |
Db    1300 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 1241

Qy     946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
             | ||  ||  |||| || |  |      ||| || | || | ||| |      |   |
Db    1240 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 1181

Qy    1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
           ||||||| ||| |  | |    |   || |   | ||   |  |  | | |||| | ||
Db    1180 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 1121

Qy    1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
           |    || |    || ||   |    | |||  ||| | || | |    ||   |
Db    1120 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 1061

Qy    1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
             || | |    || ||| | | || | | ||  | | |     || |  || |   |
Db    1060 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 1001

Qy    1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
           ||| ||| ||| ||      ||  |       ||    |||| |||  | |  |  |||
Db    1000 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 942

Qy    1245 accaactggatcattcgatgtagatatttttcttaca 1281
           |  |||   |||  |   | |   |||||||| ||| |
Db     941 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 905
```

```
RESULT  12
US-08-544-332-8/c
; Sequence 8, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-544-332-8


  Query Match             3.2%;  Score 57;  DB 2;  Length 1511;
  Best Local Similarity   45.9%;  Pred. No. 0.0023;

  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy     706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
           || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |     |
Db    1480 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 1421

Qy     766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                   || |  || | | ||| |    |||  || | |  ||     |   ||
Db    1420 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 1361

Qy     826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
             |  | ||   | ||      ||     ||   | |   | ||    |    | ||
Db    1360 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 1301

Qy     886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
             |   |||  || |   |||  ||||   | |  | ||    | || | ||     | |
Db    1300 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 1241

Qy     946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
             | ||  ||  |||| || |  |      ||| || | || | ||| |      |   |
Db    1240 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 1181

Qy    1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
           ||||||| ||| |  | |    |   || |   | ||   |  |  | | ||||  | ||
Db    1180 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 1121

Qy    1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
           |    || |    || ||   |    | |||  ||| | || | |    ||  |
Db    1120 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 1061

Qy    1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
             || | |    || ||| | | || | | | ||  | | | |    || |  || |  |
Db    1060 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 1001

Qy    1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
           ||| ||| ||| ||      ||  |       ||    |||| ||||  | |  |  |||
Db    1000 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 942

Qy    1245 accaactggatcattcgatgtagatatttttcttaca 1281
           |  |||   |||  |   | |   ||||||| ||| |
Db     941 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 905


RESULT  13
US-08-852-629-11/c
; Sequence 11, Application US/08852629
; Patent No. 6106825
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W
;    APPLICANT:  Li, Yi
;    APPLICANT:  Hall, Richard L
;    TITLE OF INVENTION:  ENTOMOPOXVIRUS-VERTEBRATE GENE DELIVERY
;    TITLE OF INVENTION:  VECTOR AND METHOD
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,629
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H
;      REGISTRATION NUMBER:  35,746
```

```
;      REFERENCE/DOCKET NUMBER:  UF-184
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  407-426-7500
;      TELEFAX:  407-839-8589
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4810 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-852-629-11


  Query Match             3.2%;  Score 57;  DB 3;  Length 4810;
  Best Local Similarity   45.9%;  Pred. No. 0.0029;
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy    706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
          || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |      |
Db   4773 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 4714

Qy    766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                  || |  || | | ||| |    |||  || | |  ||     |   ||
Db   4713 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 4654

Qy    826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
            |  | ||   | ||     ||     ||   | |   | ||    |    | ||
Db   4653 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 4594

Qy    886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
            |   |||  || |   |||  ||||   | |  | ||    | || | ||     | |
Db   4593 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 4534

Qy    946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
            | ||  ||  |||| || |  |      ||| || | || | ||| |      |   |
Db   4533 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 4474

Qy   1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
          |||||||  ||| |  | |     |   || |   | ||    |  |  | | |||| | ||
Db   4473 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 4414

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
          |    || |    || ||    |    | |||  ||| | || | |    ||   |
Db   4413 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 4354

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
            || | |    || ||| | || | | | ||  | | | |     || |  || | |
Db   4353 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 4294

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
          ||| ||| ||| ||      ||  |       ||    |||| ||||  | |  |  |||
Db   4293 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 4235

Qy   1245 accaactggatcattcgatgtagatatttttcttaca 1281
          |  |||   |||  |   | |   |||||||  ||| |
Db   4234 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 4198


RESULT  14
US-08-852-629-15/c
; Sequence 15, Application US/08852629
; Patent No. 6106825
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W
;    APPLICANT:  Li, Yi
;    APPLICANT:  Hall, Richard L
;    TITLE OF INVENTION:  ENTOMOPOXVIRUS-VERTEBRATE GENE DELIVERY
;    TITLE OF INVENTION:  VECTOR AND METHOD
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,629
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF-184
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  407-426-7500
;      TELEFAX:  407-839-8589
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4838 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-852-629-15


  Query Match             3.2%;  Score 57;  DB 3;  Length 4838;
  Best Local Similarity   45.9%;  Pred. No. 0.0029;
  Matches  265;  Conservative    0;  Mismatches  310;  Indels    2;  Gaps    2;

Qy    706 ttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaact 765
          || ||  ||||| | |||||    ||| ||    ||| |  ||| |   | | |      |
Db   4801 TTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTT 4742

Qy    766 attccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtt 825
                  || |  || | | ||| |    |||  || | |  ||     |   ||
Db   4741 GGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTT 4682

Qy    826 tcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtct 885
            |  | ||   | ||     ||     ||   | |   | ||    |    | ||
Db   4681 AAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAAC 4622

Qy    886 gaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaact 945
            |   |||  || |   |||  ||||   | |  | ||    | || | ||     | |
Db   4621 TCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATT 4562

Qy    946 tttattaatatttttaatacagcaaaggagaaaataga-tcttgtatttattcaaacata 1004
            | ||  ||  |||| || |  |      ||| || | || | ||| |      |   |
Db   4561 AATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCA 4502

Qy   1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
          |||||||  ||| |  | |     |   || |   | ||    |  |  | | |||| | ||
Db   4501 CAATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTA 4442

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
          |    || |    || ||    |    | |||  ||| | || | |    ||   |
Db   4441 TAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATA 4382

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
            || | |    || ||| | || | | | ||  | | | |     || |  || | |
Db   4381 TTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTA 4322

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaagaacaatt 1244
          ||| ||| ||| ||      ||  |       ||    |||| ||||  | |  |  |||
Db   4321 TCATAAT-ATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATT 4263
```

```
Qy   1245 accaactggatcattcgatgtagatatttttcttaca 1281
          |  |||   |||  |   | |   |||||| ||| |
Db   4262 AATAACATTATCTCTGTTTATTATTATTTTTTTTAAA 4226


RESULT  15
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             3.1%;  Score 54.4;  DB 1;  Length 5852;
  Best Local Similarity   48.4%;  Pred. No. 0.01;
  Matches  212;  Conservative    0;  Mismatches  221;  Indels    5;  Gaps    2;

Qy    846 taaagatgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaaataa 905
          | ||| | |  ||  |  ||| |  | ||  |  || | ||||    || ||| ||| ||
Db   5399 TGAAGTTCATCAAGATATATAGATAATTATTTAATTATTTGAATTTTTAAAAAAAAAAAA 5458

Qy    906 agaattacatagggcagcagcgatgttatcagcaggaacttttattaatatttttaatac 965
          | ||  | | |    | ||   |||||    |    ||  ||| ||||| |  | |||
Db   5459 AAAAAAAAAAAAAAAATCAAATATGTTTATTGTTTTAAGATTTTTTAATCTCGTCAATGA 5518

Qy    966 agcaaaggagaaaatagatcttgtatttattcaaacatacaat---ttagaaactacaaa 1022
                ||    ||||| |||     ||||    ||| |  |  |   ||    | |  ||
Db   5519 TTTTAAAATAAAAATCGATACATAATTTTAAAAAAAACCCTTTACATTTTTTATTTTAAT 5578

Qy   1023 tccagatataatggtagatatgtacttatcccatttatattttggtttaaaatataacat 1082
          ||||  || ||    |   |   |  || |    ||| | ||||  |||  ||| |||  |
Db   5579 TCCAAATTTATACATTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTTAAATT 5638

Qy   1083 cacaatcatattaggtttttcattagaacataacagaggtggatttagtcccgaaaataa 1142
            |  | ||   |||||  |||    | |    |     | ||  |    || | ||
Db   5639 TTTTTTTTTTTTTTTTTTTTATTTATTTAAAATTTAATTATTAATT--TTATAAATAAAA 5696

Qy   1143 agaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataataatagggc 1202
          |  |  | || || |   |  | ||||||  | || | | | || |||| |   | |   |
Db   5697 AATAGAAATATAAGTAAAAAAACAAACAACAAATAACATATATAAAAAAATACAAATAAC 5756

Qy   1203 agatggtatagggatatggcatttatttatgaaagaacaattaccaactggatcattcga 1262
          | ||  |  |  |  |          |||  || | | | | |  || |  |    ||
Db   5757 AAATAATTAAATAAATTAAATAACCATTAAAAATGTATATTAATAAATTTAAAAGATCTT 5816

Qy   1263 tgtagatatttttcttac 1280
          | |   || | || ||||
Db   5817 TATTAGTACTATTGTTAC 5834


Search completed: March 21, 2002, 15:41:06
Job time: 4352 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:24:29 ; Search time 1543.83 Seconds
(without alignments)
12278.265 Million cell updates/sec

Title: US-09-579-383-2
Perfect score: 1764
Sequence: 1 atgaattttaaaatatcaat.....tggatttaccattacaataa 1764

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	104.4	5.9	1101	13	CNS0021J	AL061936 Drosophil
	2	89.8	5.1	1101	13	CNS00EVL	AL069706 Drosophil
c	3	89.2	5.1	1101	13	CNS0039G	AL063921 Drosophil
c	4	84.6	4.8	1101	13	CNS00EVL	AL069706 Drosophil
	5	83	4.7	1101	13	CNS00LT2	AL078714 Drosophil
c	6	82.8	4.7	928	13	CNS0102F	AL098433 Drosophil
	7	81.2	4.6	945	13	CNS04D0K	AL285149 Tetraodon
	8	79.8	4.5	994	13	CNS04NOJ	AL298972 Tetraodon
	9	79.4	4.5	1101	13	CNS00KB5	AL077453 Drosophil
c	10	79.4	4.5	1225	13	CNS0161D	AL106171 Drosophil
c	11	79.2	4.5	1161	13	CNS073Y8	AL428118 clone BA0
	12	78.4	4.4	1101	13	CNS0006J	AL062049 Drosophil
c	13	77.6	4.4	524	13	CNS01U9O	AL167541 Tetraodon
	14	77.4	4.4	876	13	CNS009G1	AL053529 Drosophil
c	15	77.2	4.4	839	11	BF259858	BF259858 HVSMEf002
	16	77.2	4.4	1190	13	CNS02ON7	AL206908 Tetraodon
c	17	77	4.4	879	13	CNS01JRG	AL147405 Anopheles
	18	77	4.4	963	13	CNS00A4L	AL054918 Drosophil
	19	76.8	4.4	1007	13	CNS06X9S	AL419462 T3 end of
	20	76.8	4.4	1101	13	CNS00EO7	AL069440 Drosophil
c	21	76.6	4.3	942	13	CNS018GS	AL109318 Drosophil
c	22	76.4	4.3	1201	13	CNS0167M	AL106396 Drosophil
	23	76.2	4.3	1101	13	CNS003B4	AL064082 Drosophil
c	24	76	4.3	1337	5	BF630719	Bf630719 HVSMEb001
	25	75.6	4.3	928	13	CNS00DKY	AL071865 Drosophil
	26	75.2	4.3	1091	13	CNS014AC	AL103902 Drosophil
c	27	75	4.3	816	13	CNS02KJV	AL201604 Tetraodon
c	28	74.8	4.2	1009	13	CNS00587	AL057153 Drosophil
c	29	74.6	4.2	878	10	AL536095	AL536095 AL536095
	30	74.6	4.2	971	13	CNS06D25	AL393267 T3 end of
	31	74.6	4.2	1101	13	CNS003BD	AL064091 Drosophil
	32	74.6	4.2	1125	10	AL547503	AL547503 AL547503
	33	73.4	4.2	1101	13	CNS0039G	AL063921 Drosophil
	34	73.4	4.2	1101	13	CNS003B0	AL064078 Drosophil
	35	73.4	4.2	1101	13	CNS00LJT	AL068307 Drosophil
	36	73.2	4.1	829	13	CNS03LUB	AL249932 Tetraodon
	37	73.2	4.1	1146	13	CNS021G2	AL176843 Tetraodon
c	38	73	4.1	1101	13	CNS017V2	AL108536 Drosophil
	39	73	4.1	1151	11	BG309087	BG309087 HVSMEc000
c	40	72.8	4.1	839	13	CNS011KX	AL100395 Drosophil
	41	72.8	4.1	1101	13	CNS0182P	AL108811 Drosophil
	42	72.8	4.1	1101	13	CNS058GV	AL325912 Tetraodon
c	43	72.6	4.1	1101	13	CNS00KAE	AL077628 Drosophil
	44	72.4	4.1	961	13	AZ691838	AZ691838 ENTMI56TF
c	45	72.2	4.1	922	13	AZ548363	AZ548363 ENTDZ07TF

ALIGNMENTS

RESULT 1
CNS0021J
LOCUS CNS0021J 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL061936
VERSION AL061936.1 GI:4940214
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR05N11"
                     /note="end : TET3"
BASE COUNT      631 a      7 c     28 g    289 t    146 others
ORIGIN
```

```
  Query Match             5.9%;  Score 104.4;  DB 13;  Length 1101;
  Best Local Similarity   40.4%;  Pred. No. 2.5e-06;
  Matches  360;  Conservative   69;  Mismatches  460;  Indels    2;  Gaps    2;

Qy      70 aaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaataaaact 129
           |||  |||||| || | | | ||   |     || || |    |||| ||||| ||||
Db     207 AAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 266

Qy     130 catcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgta 189
           |  |||   ||| | |  |          | |    |||   |: || |   |      |
Db     267 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAAAAAAACAAAAAAAA 326

Qy     190 gaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaaaat 249
           || |   |    |      | |        ||    | ||  | |||||| || |||||
Db     327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 386

Qy     250 ataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaaggaaa 309
           | ||| |||||  | | ||||  |  || | |||    | |  |  | |||||||  |||
Db     387 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 446

Qy     310 caaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacat 369
           ||   |  | |  |  | |  |      |   || |   ||    | | |  |||| |
Db     447 AAAAAAAAAAAAAAAAAAAANAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 506

Qy     370 atgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatg 429
           |  |   ||  ||    : |  | :  | ||:| | ||   |||  |    :| |  |
Db     507 AANAAAAATATAATTTAWTTTTTTWTTAATTAWTTTTTTTTTTTTTTTTTTTWTTAATTT 566

Qy     430 ttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggc 489
           | || |    :|:  |    ||  ||:   :   |:   |   :   :: |||| |
Db     567 TAATTTTTAAWAWAAATTTAATAAAAWAWTWTTTAWTTTTAATWTAAWWAAAAAAAAWTT 626

Qy     490 ttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgt 549
           ||| |: :|      | |  ||      :||  ||: || |  | :   | :: |:
Db     627 TTAAAWAWTWTTTTTTTTWTTAATTAATAWAAAAAAWATAWAAWTTWTTTATWWATWAAA 686

Qy     550 ccagatgtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaaa 609
           | || | |  | : ||||   ||   |    ||| | ||| |:   || | :| :|
Db     687 TWAAATTTTAAAAAWTTTTTTAWTTTTTTTWTTTAAAAWTTTATTAWTTTTAAAAWATWA 746

Qy     610 gaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgta 669
           ::|  ||| | |  |||: | : | |:     ||  | ||||   |||   |  |  |
Db     747 WTWWTAAATTTTTTAWATAWTAAWATTAWAAAATTTTTTAATTTAATTTTTTTTAATTAA 806

Qy     670 gatattgactgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaaat 729
           | |  |     ||  | ||   ||   | | ||   |||: ||::   ||||  |
Db     807 AAAAAAAAAAAAAAAAWAAAT-AAAAAAATAAAAAAAAAAAWAAAWWTTTTTTTAAATTA 865

Qy     730 tattatattaaattaattaacttgttaagaaaaactattccggaagaaaagttaatttca 789
           | |: |||| ||  | || | |  | |   ||:  | ||    ||| : |   |:| : |
Db     866 TTTWTTATTTAAAAATTTTAATAATSATTTAAWTATGTTTTTTAAGTWTAAAAAWTGKGA 925

Qy     790 atttctggttcatcaaatgctgcattatcatgcgtttcaggagttgcatctttctgtaaa 849
           || : ::|:| ::   ::| |  |  ||  |  | :      :|:  |  ::|    ||:
Db     926 ATGKKKKGKTARKKKGRKGTTA-ARAATDGTAAGAKRATAAKKTKTKAATKKTAAAAAAD 984

Qy     850 gatgaagaatctccatataacactaaatttttgtctgaacaaatagaaacaaataaagaa 909
           :|:||| ||| :  |  :|  |  || | ::  | |     || ::||| ||| ||| ||
Db     985 KAKGAAAAATAWGAARGKATGAAAAAGTGKDAATATTTKKNAAAWKAAAAAAAAAAAAAA 1044

Qy     910 ttacatagggcagcagcgatgttatcagcaggaacttttattaatattttt 960
           | | || | |:     |:   |: |:   |   :   : ||:|  :|:|
```

```
Db    1045 AAAAAAAGAGAAKAKARAAKAAAADAAKAGAGTTTDAAADATAWTKGKTKT 1095


RESULT   2
CNS00EVL
LOCUS       CNS00EVL     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN
```

```
  Query Match             5.1%;  Score 89.8;  DB 13;  Length 1101;
  Best Local Similarity   34.8%;  Pred. No. 0.00032;
  Matches  218;  Conservative  108;  Mismatches  296;  Indels    5;  Gaps    1;

Qy     716 taaatttttcaaattattatattaaattaattaacttgttaagaaaaactattccggaag 775
           | |:|:|  :::::::| |:|:: ||:  |:|||     : || |:|:: :|||    ::
Db     457 TCAHTWTMMMMMWWAATWTWWAAAWAAAWTTATWAATWAAAAAAWAWWWWATTTTTWWWW 516

Qy     776 aaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgtttcaggagttg 835
           : :| |:  |:  : :|  |     |  |:||  |  |  :: ||   ||  :    |
Db     517 WTWATTWTTTWAWWTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAW 576

Qy     836 catctttctgtaaagatgaagaatctccatataacactaaatttttgtctgaacaaatag 895
           |: | | |  | : || || : : |  ||||||  |  | ||||||  : | ||    |
Db     577 AAWWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTA 636

Qy     896 aaacaaataaagaattacatagggcagcagcgatgttatcagcaggaactttattaata 955
           | |     |   |||||  ||    |  |  || :::| |  |   | |||:|   |:|
Db     637 ATA-----ATTTAATTAWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWA 691

Qy     956 tttttaatacagcaaaggagaaaatagatcttgtatttattcaaacatacaatttagaaa 1015
           || | |  |:|  | :|| :| |:  |  : |  | : | | | : |:  |: ::|
Db     692 AAAAAAAAAAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWA 751
```

```
Qy   1016 ctacaaatccagatataatggtagatatgtacttatcccatttatattttggtttaaaat 1075
          | | ::  | |||:  :  :: ||:: :: |      |:    :  :   : ||||
Db    752 TAATATWWATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATA 811

Qy   1076 ataacatcacaatcatattaggtttttcattagaacataacagaggtggatttagtcccg 1135
          :  | || | |:  |:|| :  |: |  |::|  | |:|| | |  |     |:| |
Db    812 WATAWATWAAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATA 871

Qy   1136 aaaataaagaattattagaattggtaggaaaaacaatacatgataaaaatcaaaataata 1195
          |:|| ||| |: : |:: : ||  :|  | | | |: :    | |:||| ||||:||
Db    872 AWAAAAAATAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAT 931

Qy   1196 atagggcagatggtatagggatatggcatttatttatgaaagaacaattaccaactggat 1255
          | :      ||  |:|    | |:  ||::|||:|| ::|  : |  :    | :
Db    932 AAWAWWWTWATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATA 991

Qy   1256 cattcgatgtagatatttttcttacaaatatttggaaacatttaaatcctgaagtacaaa 1315
          ||:  |:  :     |||  | | : ||||||  :|| :|   |      |: :  :|
Db    992 TATWWTAWTAWATATATTTATTAAWWTATATTTTAWAAAWTAATATATWATAWWWTAWAT 1051

Qy   1316 ctccaaaagaccttactataactgaaa 1342
          |  |::|    :|  ||||  | |:|
Db   1052 ATAWAWWAATTAWTTATATATWTAAWA 1078


RESULT  3
CNS0039G/c
LOCUS       CNS0039G     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR08K10"
                      /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match             5.1%;  Score 89.2;  DB 13;  Length 1101;
  Best Local Similarity   19.3%;  Pred. No. 0.00039;

  Matches  136;  Conservative  302;  Mismatches  263;  Indels    5;  Gaps    2;

Qy     18 aatatttttaattatagtatccatcttgtattctgcaaattccagaaccttgaaaggaaa 77
          :: ::|:: :  :::::: :   : ::: ::  :  :|: :  :: |  :: | :  :::
Db   1097 RWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWW 1038

Qy     78 aaataatataaataattcattgggaataatacgggaaaataaaaataaaactcatcaaac 137
          ::||::|::: :::: | |:|:  : :::|| : ::| ::: |   :::| | |: | :
Db   1037 WWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRR 978

Qy    138 ggaaatacatgagtctttttcacatcttaaat--cgaataatagtaattttgtagaatat 195
          :| :     ::::: :::::    :    ||:   :::::::::::::::  ::: |:::
Db    977 DGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDD 918

Qy    196 ggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaaaatataaat 255
          :||  :::: : ::|::: :||:::: : :::::::|   : :||:|:|| ::::: :|:
Db    917 DGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWAD 858

Qy    256 aaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaaaggaaacaaggt 315
          |::: ::| |::::|: :  |: : ::: : :  :  :: :: ::|  :|||:  ::::
Db    857 ADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKR 798

Qy    316 attatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacatatgatt 375
          |::  :  :::::| |:|  ||  |:: : :::  :  |:: |::  :::: :::::| |
Db    797 ADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGT 738

Qy    376 gattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatgttatat 435
          : ::    ::::::   :: : ::|| :: : : ::| : |  ::     :  : ||
Db    737 AGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRAT 678

Qy    436 gatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggcttagaa 495
          :  :|:  : |   ||: :|:::::: ::::      :|||: |:::   :  ::|:::
Db    677 WDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWD 618

Qy    496 tatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagat 555
          : |:|::   :  :|:|:|: : : :: :|:|: |:: : :  : :::: |  :|  | :|
Db    617 WKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWT 558

Qy    556 gtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaaagaaatt 615
          :||  |:: :: :||   |:  |    :    | | |  |  :| || :::|| : ||:
Db    557 TWAARAAWWAWWAWATTTATWTTTWTTTW---TTTTTTTTTTTWTTTAAWWWAAWWTATW 501

Qy    616 gattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgtagatatt 675
          |:||  :  | |||: |||     || ||:|: |  :: |:|:   |  | |: |  ||
Db    500 AAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTT 441

Qy    676 gactgggaaccacatgggaagttttacaacttaaatgaattaaatt 721
          :    :  |       : ||:|| ||   |::: : :|| ||:
Db    440 TTTWTWTTWWAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATATK 395


RESULT  4
CNS00EVL/c
LOCUS       CNS00EVL     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
```

```
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
BASE COUNT      419 a      91 c      60 g     299 t     232 others
ORIGIN


  Query Match             4.8%;  Score 84.6;  DB 13;  Length 1101;
  Best Local Similarity   32.9%;  Pred. No. 0.0018;
  Matches  205;  Conservative  118;  Mismatches  297;  Indels    3;  Gaps    1;

Qy     142 atacatgagtctttttcacatcttaaatcgaataatagtaattttgtagaatatggatct 201
           ||: :|   | |||:: | :|  :| ||  ||: |   :: :| | |: |:::|  :
Db    1096 ATWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTATWATA 1037

Qy     202 tattgcggagatgggtgtaactctagaattacaaaaaataataaaaatataaataaaaat 261
           ||||          | ||: |  | || || | : |:||::| | |||:| |:||: ||
Db    1036 TATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTAT 977

Qy     262 gatagaaaatcaccaagacaaattttagaggagtataaaaaaaggaaacaaggtattata 321
           : : ::||| |  ::   :|:||| | |  | |||:|::: :  |       | || |:
Db     976 AWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTW 917

Qy     322 gcaggatactatggttcatggaacagtcaaggtgatagagcaaaacatatgattgattca 381
           ||: : |  |  ::  :| | : ::  : :||     :  ||:   :  ||| :
Db     916 TTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAW 857

Qy     382 aacccaatggtgtcaattttatatattgcatttgctcgcattaatatgttatatgatgtta 441
           |      :  | | ::| |  :||::|  :|:|  |    :|| :   ||||:|  : |:
Db     856 ATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTW 797

Qy     442 tctagaccatttaatgga---agacaaagattcctattaagaaaacacggcttagaatat 498
           | |     :| ||::  :   : : |:| :|   |||::| |  | :    ||:  :| |
Db     796 TATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTT 737

Qy     499 gaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccagatgta 558
           |    | |  : |  :  |:| |   : |:     | |          :   |  | ||
Db     736 TATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTAT 677

Qy     559 attattcttttatccttaggtggagaaacctatatgatagatatagaaaaagaaattgat 618
           | ||| ::: ||  :|   |  | ||:    ||  ||   | |     | :||   ||
Db     676 TTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAAT 617

Qy     619 tatgtggataaaatattgaagcttgttaatgattttgatttagatggtgtagatattgac 678
           | |      |   ||: :|  |   :| | || |  :||:  |   : ||  :  | :
Db     616 ATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWW 557

Qy     679 tgggaaccacatgggaagttttacaacttaaatgaattaaatttttcaaattattatatt 738
           |   ||  | :|      || :  :| :|:||  |||:|:::::   || ::::|:| ||
Db     556 TTTAAATTATWTTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTT 497

Qy     739 aaattaattaacttgttaagaaa 761
           :|||:|| |: || || : |:|
Db     496 TWATTWATAAWTTTWTTTWWAWA 474
```

```
RESULT   5
CNS00LT2
LOCUS       CNS00LT2     1101 bp    DNA             GSS     14-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078714
VERSION     AL078714.1  GI:5102004
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR48P19"
                     /note="end : TET3"
BASE COUNT      469 a      6 c      69 g     151 t     406 others
ORIGIN


  Query Match             4.7%;  Score 83;  DB 13;  Length 1101;
  Best Local Similarity   35.2%;  Pred. No. 0.0031;
  Matches  222;  Conservative  102;  Mismatches  306;  Indels    1;  Gaps    1;

Qy      70 aaaggaaaaaatataataaataattcattgggaataatacgggaaaataaaaataaaact 129
           |||  |||||| || | ||| ||   |     || || |    |||| ||||| ||||
Db     225 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 284

Qy     130 catcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgta 189
           |  |||   ||| | |  |          | |    |||   || || |  ||      |
Db     285 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 344

Qy     190 gaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaaaat 249
           || |   |    |      | |       ||    | ||  | |||||| || |||||
Db     345 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404

Qy     250 ataaataaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaaaggaaa 309
           | ||| |||||  | | ||||  |  || | |||    | |  |  | |||||||  |||
Db     405 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 464

Qy     310 caaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacat 369
           ||   |  | |  |  | |  ||  ||  |  :  | |  :  | : |::: ||::  |
Db     465 AAAAAAAAANAAAAAAANAAAAATTWTTTTTTTWKKATTTTWTTTTDAADDKWAAWWDTT 524

Qy     370 atgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatg 429
           | |:| :|| |     | |  | |  :||::|| |::|: |:|   |   | :: | |
Db     525 TTTAWTTWTTWATTAATATTTTTTTTKWTTWKATKTDWTKKAKTAAKTAAWAAWWTTTTT 584
```

```
Qy    430 ttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggc 489
          | ||:|  |:|:| ||:   |::::|:: |      :  ||  | || : |
Db    585 TAATWTTTTKTWT-TAKKTWAWKWWAWDTATTTWTTWTTTTTWTTTTTWAATTTTTWTTT 643

Qy    490 ttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgt 549
          :|| | | |        |: | |  | |     | |  | |  | : : ||:: |  ::
Db    644 WTAWATTTTTTTTTTAKTKTTTTTTTTTTATTAAATAAAAAAWTTTDTWAAAWWTTTKKK 703

Qy    550 ccagatgtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaaa 609
            | : :::|:: :  :    |: :::: | |||  ::  :| : : : : |
Db    704 KKKAAAADKWKDAKKWDGAKKATKTKKKDKKAAWAAADKKDRKKGGKKKGKGKKGKKGGK 763

Qy    610 gaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgta 669
          :   ::: :  :|::| : | ::: ::  : :::: | |  ::::| ::: :| :::: |
Db    764 KKKKKKKGKGGWGKKGKKKAGDDDAKDKTKKKKKKAATTTTKKKKGKKKKKKAAKKKKAA 823

Qy    670 gatattgactgggaaccacatgggaagtttt 700
          ::| :|::    ||     :::|  :::|:
Db    824 DRTKKTKWDAAAAAAAAKKTKDKGKKKKKTK 854


RESULT   6
CNS0102F/c
LOCUS       CNS0102F      928 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03B09 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098433
VERSION     AL098433.1  GI:5610044
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03B09"
                     /note="end : T7"
BASE COUNT      169 a    128 c    100 g    241 t    290 others
ORIGIN


  Query Match             4.7%;  Score 82.8;  DB 13;  Length 928;
  Best Local Similarity   26.8%;  Pred. No. 0.0035;
  Matches  136;  Conservative  160;  Mismatches  211;  Indels    1;  Gaps    1;

Qy    119 aaaataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaataata 178
          || |:: :| : :  :|| :::|  :   ::::   | :      :|| :  ||  :| :
Db    923 AAGAKWGWAWWGWRRWAATDRWATAWTKAKWRWGAGTRWATAKARKTAKRGAGAKRWAAR 864

Qy    179 gtaattttgtagaatatggatcttattgcggagatgggtgtaactctagaattacaaaaa 238
            |: ::||:  |::|: :: :|  || ||:  | :|         || : |: :: :
Db    863 WATAKAKWGTRWGAKWTRADKKDTGGTGTGGWTGTWRGRTGKKRRGRAGGDGTRGDRTRT 804

Qy    239 ataataaaaatataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtata 298
          : | :| ::|:| :| ||:||::|| : :|:|  :   |||  |::: | : |  |:
Db    803 RRAGKAGRWAKAARAKTAWAADWGARWAWAWAAWDATGAGATAWWWATAWTATAAKATAT 744

Qy    299 aaaaaaggaaacaaggtattatagcaggatactatggttcatggaacagtca-aggtgat 357
            |:|:: :::| || || :::| |   :| | :   ::| |: : |  || | |:   ||
Db    743 TAWAWRWRRDARAATGTGKWRTGGTTAKAGATKGGWDDTRADAKKAAGGTTATAKTATAT 684

Qy    358 agagcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatttgct 417
            |: :|:| |::::|| :|:  | |   |:  | :  ||:| :: ||  :   | | |
Db    683 GTAKDRARAKAKWWKATAKAKAGATATRAAWATGADAWGATWTGWKKTAGAKTTGTAGWT 624

Qy    418 cgcattaatatgttatatgatgtatctagaccatttaatggaagacaaagattcctatta 477
            : :|: : :: ::::|||::: || :|||  : : : ::   || ::: : :   | :
Db    623 GRKWTDKWRRKTDKRDATGRWRAATAWAGAADDAWRDGWRAGTGAGRRRWDGWRWRARAD 564

Qy    478 agaaaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatc 537
          : : |  | ::   | ||  ||  |   :::  : : :|  :: :   :   |  : |
Db    563 RTRKAGAADRRAGAAAAAAGTGTGADAADKRAARGKKDGDGKRKKTGKRAATGGWRGAGR 504

Qy    538 agaaaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatatgata 597
          |:: : : :  :  ::|:::  |:::| |::     | |:: :||:::|  :::  |  |
Db    503 AKDGDKRGWGWKKKRKAKDKTKTKWDTRTDKAKAKATGAKDGDGAKRWAGDKRKKGGGGA 444

Qy    598 gatatagaaaaagaaattgattatgtgg 625
          |||||:  : :::::|||||:|:| ||:
Db    443 GATATRTGWGWWDWWATTGAWTWTATGR 416


RESULT   7
CNS04D0K
LOCUS       CNS04D0K      945 bp    DNA             GSS       21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            101H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL285149
VERSION     AL285149.1  GI:8023560
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 945)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
```

```
                     /note="Genoscope sequence ID : C0BG101CD11LP1~end : T7"
BASE COUNT       386 a    112 c     96 g    231 t    120 others
ORIGIN
```

```
  Query Match             4.6%;  Score 81.2;  DB 13;  Length 945;
  Best Local Similarity   41.8%;  Pred. No. 0.006;
  Matches  212;  Conservative   52;  Mismatches  240;  Indels    3;  Gaps    2;

Qy     703 aacttaaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaa 762
           :::||:  |   ||  | |  :   |:::||:| ||   |  | | | |    :  |||
Db     429 MMHTTWNTTTTTTTTTANTANYWTTAWWWTTWTTTTTTTTTATAATTAATAAAAWTTAAAN 488

Qy     763 actattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgc 822
           : | :|    |: :|||   ||  | |: :: |  || |  |||:     | :|| :|
Db     489 WATTWTWAAAAWAWAAATAAAAAATAAWAWWWTTTTTAAAAAAAWATAATAAWATTWTWT 548

Qy     823 gtttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttg 882
           : : |    :    | ||| | |||: || |   |      | ||| | :|||| |||
Db     549 AWAWWATTTTWATATTTTTTATTTAAWAATAATATAAAAAATTTTAAAAAWAAATATTTT 608

Qy     883 tctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcagga 942
           |   || |||:| ||| | | | | | ||| |:|    |  :   :: | :: |  :  :
Db     609 TAAWAATAAAWATAAAAATAAATATATTTATAWAAWAWATAWAWWWWWTAWWAAWWWTAW 668

Qy     943 acttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaaca 1002
           | | | |  ||||   : :|   |  | | |  | |||| : |  :   ::|:||  :|  |
Db     669 AATATAAAAAATAAAAWAWAATAAAAATAAAATAAAAAWAAAWWAAAWWTWATAAWANAA 728

Qy    1003 tacaatttagaaacta-caaatccagatataatggtagatatgtacttatcccatttata 1061
           | ||  || | | ||  |||| | ||  |||   | |||: |:  ||    |   |
Db     729 AAAAAAATATATAATAWAAAATATAAATTAAATATAAAATAWATWWATAAATAAAAAAAT 788

Qy    1062 ttttggtttaaaatataacatcacaatcatattaggtttttcattagaacataac--aga 1119
           ||:|  |   ||| |:|  |  :     :|||||  |     ||| ||  ||:   | |
Db     789 TTWTAATAATAAAAAWATTAAAWATTATWTATTANWATAAAANTTATAAWNTAWNTAATA 848

Qy    1120 ggtggatttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgat 1179
           |  | |:       |||| |:| ||||||  | || |   |  | ||| |  | || |
Db     849 AATAAAATWTAAAAAAAAAAAAWAAAATTATAAAAAATAAAATTATAAAAATAATATAAA 908

Qy    1180 aaaaatcaaaataataatagggcagat 1206
           |||||  |||| ||:  ||    | ||
Db     909 AAAAAAAAAAAAAATWTTAAAAAAAAT 935
```

```
RESULT   8
CNS04NOJ
LOCUS       CNS04NOJ      994 bp    DNA             GSS       21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            122P04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL298972
VERSION     AL298972.1  GI:8037822
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 994)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 994)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .994
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="122P04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG122DH02LP1~end : T7"
BASE COUNT       543 a     49 c     36 g    194 t    172 others
ORIGIN
```

```
  Query Match             4.5%;  Score 79.8;  DB 13;  Length 994;
  Best Local Similarity   40.4%;  Pred. No. 0.0094;
  Matches  271;  Conservative   29;  Mismatches  371;  Indels    0;  Gaps    0;

Qy      49 tctgcaaattccagaaccttgaaaggaaaaaataatataaataattcattgggaataata 108
           | :| |||    | ||     |||  |||||| || | ||| ||   |      |: | |
Db     319 TTBGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWAAAAATNAWNAAA 378

Qy     109 cgggaaaataaaaataaaactcatcaaacggaaatacatgagtctttttcacatcttaaa 168
           |||| |||||  | |   |  |||   |||:| |  : | || |  : |  : :||
Db     379 NAANAAAAAAAAAAATANAANAAAAAAAAATAAAAWAAAAAWATTTTATATWTAATWAWAA 438

Qy     169 tcgaataatagtaattttgtagaatatggatcttattgcggagatgggtgtaactctaga 228
           |   |||||| |||     || | |:      | |||    | ||     | | | ||
Db     439 TTANATAATAATAAAAAAATATATTWAATNAATAATTAATAATATNATATTTATTATATN 498

Qy     229 attacaaaaaataataaaaatataaatataaaaatgatagaaaatcaccaagacaaatttta 288
           || | :| ||| |   | |  ||||| | |||  | | || :: |     |  |:|  ||
Db     499 ATAAAWATAAAAATATATATAATAAANATAAAAAAAAAAAATWWTATATTAATNAWTAATA 558

Qy     289 gaggagtataaaaaaaggaaacaaggtattatagcaggatactatggttcatggaacagt 348
           | ||  ||| ||  ||| ||     : || |  | | |  ||   | |    |  |
Db     559 TTATATTAATAAATAAAAAAAAAAATNATAWNTAATATAAAAAAATATATAAAAATAWTTT 618

Qy     349 caaggtgatagagcaaaacatatgattgattcaaacccaatggtgtcaattttatatatt 408
           |   ||| |    |   : :  | | |  ||    |  |   | : |||  | ||
Db     619 ATAANAWATAAATATTATATWTWTTTATAATATAANAATANTATTATATWATTAATTTTT 678

Qy     409 gcatttgctcgcattaatatgttatatgatgtatctagaccatttaatggaagacaaaga 468
           ||||      |  |||    | || |:  |  | |  |   ||  || | ||| |
Db     679 AWATTTWAAAWNAAATATAAWNAAAATAAWAAAATAAAAAAAAAAAAAAAAAAAAAAAA 738

Qy     469 ttcctattaagaaaacacggcttagaatatgaaacctatggtatgatgcttaatgaaatt 528
           :|||    | | |   : | |||    |:   |    |  ||    |   |:|
Db     739 AAAAWATTTTATATATAANAAWAAAAATTAATAWAAAAAAAAAAATATATAAATATAWAAA 798

Qy     529 agacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtggagaaacc 588
           | |   |  | || | ||  |     :  ||  |          ||  | |  | :   ||
Db     799 AAAWAAAAAAAAATAATAAATTNTANWTAAAAAAAAAAAAAATAATAATATTRTTTAATA 858

Qy     589 tatatgatagatatagaaaaagaaattgattatgtggataaaatattgaagcttgttaat 648
           : |        :|| |||:: ||    |||   |    ||||:||  ||  || ||
Db     859 AWAAAATNTTTAWTAAAAAWWAAATAANATTTAATAATAAAAAWATATAATTTTTTTGTA 918

Qy     649 gattttgatttagatggtgtagatattgactgggaaccacatgggaagttttacaactta 708
           |: |   |||| |     |  || | | | | | | | ||      || |  ||
Db     919 AAWATAATTTTAAAAAAAATTAATTTAAAATTTGTATTAAATTTTTTATTATGAAAAGAT 978

Qy     709 aatgaattaaa 719
           ||  |  |||
Db     979 TATTTAAAAAA 989
```

```
RESULT   9
CNS00KB5
LOCUS       CNS00KB5     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR39P05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077453
VERSION     AL077453.1  GI:4956930
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR39P05"
                     /note="end : T7"
BASE COUNT      540 a      0 c    139 g    252 t    170 others
ORIGIN

  Query Match             4.5%;  Score 79.4;  DB 13;  Length 1101;
  Best Local Similarity   41.2%;  Pred. No. 0.01;
  Matches  282;  Conservative   28;  Mismatches  370;  Indels    5;  Gaps    1;

Qy      54 aaattccagaaccttgaaaggaaaaaatatataaataattcattgggaataatacggga 113
           |||    | ||     |||   ||||| ||    || ||   |     || || |    |
Db       5 AAAAAAAANAAAAAAAAAAAATAAAAANAANNANAAAAAAAAAAAAAAAAAAAAAAAAA 64

Qy     114 aaataaaaataaaactcatcaaacggaaatacatgagtctttttcacatcttaaatcgaa 173
           ||| ||||||||||   |  |||   ||| | |  |         | |    |||   ||
Db      65 AAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 124

Qy     174 taatagtaatttttgtagaatatggatcttattgcggagatgggtgtaactctagaattac 233
           ||||  ||      | || |   |   || |     |          || | || |||
Db     125 AAATAAAAAAAAAAAAAAAAAAAAAAAATAATAAATAATATAAAAAAATTATAAAATATT 184

Qy     234 aaaaaatataaaaatataaataaaaatgatagaaaatcaccaagacaaattttagagga 293
           |||:||:||||||||| | ::|  ||   || :|||  |  :|   |||    | | |
Db     185 TAAAWATWATAAAAATAAATWWATTAAAATTAWWAAAAWAAAWAATTAAAAAAAAWATAA 244

Qy     294 gtataaaaaaaggaaacaaggtattatagcaggatactatggttcatggaacagtcaagg 353
           | |:|||:|  :|| ||      | |  |  | |  |      |    | |   ||
Db     245 AAAAAWAAAWATAWAAAAAAAATAAAAAAAAAAAAAAAAAATAAAAAAAATAAAAAAAAT 304

Qy     354 tgatagagcaaaacatatgattgattcaaacccaatggtgtcaattttatatattgcatt 413
           | | |  |||| | |  |   |          |       |||| || | | |  ||

Db     305 ATAAAAAAAAAAAAAAAAAAAAANANNNNNNAAATAAAAAAAANAATTATANANAATTTATA 364

Qy     414 tgctcgcattaatatgttatatgatgtatctagaccatttaatggaagacaaagattcct 473
             |    ||| | | |||| | || | |  | |  ||  ||   || | ||| |
Db     365 ATATTTTTTTATTTTTTTATTTTATATTTTAAAAAAATAAAAATAAAAAAAAAA----- 419

Qy     474 attaagaaaacacggcttagaatatgaaacctatggtatgatgcttaatgaaattagacg 533
           | ||| |||| |      | || |:  ||   |   |  |   | ||  |||  |
Db     420 AATAAAAAAAAATAAAAAAAAAAAAWATAAAAAAAAAAAAAAAAAAATNAAANAAANAANNNN 479

Qy     534 tatcagaaaagtacgtccagatgtaattattcttttatccttaggtggagaaacctatat 593
           |  |  |||     |  | |   |||| || || ||     |     | |||   | |
Db     480 AAAAAANAAATATAATTTAAAAAAAATTTTTTTTNTAAAAAAAAAAAAAAAAAAAAAAAA 539

Qy     594 gatagatatagaaaaagaaattgattatgtggataaaatattgaagcttgttaatgattt 653
           | | | | | | ||| ||    |  |     | |||| |   | :  || : :: : ::
Db     540 AAAAAAAAAAAANAAATAANAAAAAAAANANAAANAAAAAAAAANAGKGATGAKGRKARRKW 599

Qy     654 tgatttagatggtgtagatattgactgggaaccacatgggaagttttacaacttaaatga 713
           | |  |  |::|:  || | |    ||     :  |     ||  |:    :|   :|
Db     600 TAAAATTTKTKRTKGKGAAAATTTKGGGTTGTTWTTTWTATKGTGGTWGGGKWTTTKKGG 659

Qy     714 attaaatttttcaaattattatatt 738
             :  | :|:     || ||||||:
Db     660 KADTTAGKTKGTGTTTTTTTATATK 684


RESULT  10
CNS0161D/c
LOCUS       CNS0161D     1225 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15C18"
                     /note="end : SP6"
BASE COUNT      266 a    128 c     38 g    368 t    425 others
ORIGIN

  Query Match             4.5%;  Score 79.4;  DB 13;  Length 1225;
  Best Local Similarity   29.8%;  Pred. No. 0.0099;
  Matches  177;  Conservative  139;  Mismatches  277;  Indels    1;  Gaps    1;

Qy     646 aatgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttacaac 705
           ::| : |:  |||| :|: | |  ::|:  | :   :|  | |:   :  | | | ||
```

```
Db   1224 WWTTWATWATATTTTTWTKATTTTTWWAWAWAWWWWAWAWAAAAWWTTWTTTATAAAAAT 1165

Qy    706 ttaaatgaattaaatttttcaaa-ttattatattaaattaattaacttgttaagaaaaac 764
Db   1164 ATATAWATWTATTATWWWAAAWAWTTTTTATATWATTWTAWATAAAAWAAAAATATTAWW 1105

Qy    765 tattccggaagaaaagttaatttcaatttctggttcatcaaatgctgcattatcatgcgt 824
Db   1104 WAWWWAAWAAWAWAWAAWWATWWTWTWKAATATAAAARTAWATATATAAAAATWAWAWWW 1045

Qy    825 ttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgtc 884
Db   1044 TWAWAAAATATATTTTTTTTTWTTTTAWAAAAAAAAWAAAWATAAATTATWWATAAAWWW 985

Qy    885 tgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaac 944
Db    984 TAAAAAAAAATAAWTTATTTWAWAWTTTTTTRAAAWWTWTTTTTTTTTWWTTTAAWAAWT 925

Qy    945 ttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaacata 1004
Db    924 WWTTATWTRTTTWTRTRRRARAATWWWTTTTTTTTTTTTTTTATATTTTTTTTTMYHHT 865

Qy   1005 caatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatattt 1064
Db    864 YTYMTTHMHHMMMMMMMHMMMGGBGGKGGGTTTTTKKTTTKTTKTKTTTGMGATTTTA 805

Qy   1065 tggtttaaaatataacatcacaatcatattaggtttttcattagaacataacagaggtgg 1124
Db    804 AATTTTTNRTDTTWARAAAWDDTTWNKTKTMTTHTTKDMKMKMGWMNNMMMVMGGMTRTK 745

Qy   1125 atttagtcccgaaaataaagaattattagaattggtaggaaaaacaatacatgataaaaa 1184
Db    744 TKTTKKDGRRRMGGWTDTDARAAAAAAAAAGADARAAARARAARGAGAGGGGGAGGGRRG 685

Qy   1185 tcaaaataataatagggcagatggtatagggatatggcatttatttatgaaaga 1238
Db    684 GARGGAARRDWAKAANGAWKRTAWTRRDARRKGGAGGTTAATTTTTTTAWAARM 631


RESULT  11
CNS073Y8/c
LOCUS       CNS073Y8     1161 bp    DNA             GSS       07-JUL-2001
DEFINITION  clone BA0AB023G11 of library BA0AB from strain CLIB 210 of
            Kluyveromyces lactis, genomic survey sequence.
ACCESSION   AL428118
VERSION     AL428118.1  GI:12211312
KEYWORDS    GSS.
SOURCE      Kluyveromyces lactis.
  ORGANISM  Kluyveromyces lactis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
            Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 11.
            Kluyveromyces lactis
  JOURNAL   FEBS Lett. 487 (1), 66-70 (2000)
  MEDLINE   20584721
REFERENCE   2  (bases 1 to 1161)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
REFERENCE   3  (bases 1 to 1161)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Kluyveromyces lactis"
                     /strain="CLIB 210"
                     /variety="lactis"
                     /db_xref="taxon:28985"
                     /clone="BA0AB023G11"
                     /clone_lib="BA0AB"
BASE COUNT      379 a      32 c      81 g     508 t     161 others
ORIGIN


  Query Match             4.5%;  Score 79.2;  DB 13;  Length 1161;
  Best Local Similarity   37.4%;  Pred. No. 0.011;
  Matches  386;  Conservative   74;  Mismatches  566;  Indels    6;  Gaps    2;

Qy      50 ctgcaaattccagaaccttgaaaggaaaaaatatataaataattcattgggaataatac 109
Db    1127 YTAAAAAAAAAAAAAAAAAAAAAAAAAAATCAAAAAAWAAAAAWAAAAAAAAAAAAATAA 1068

Qy     110 gggaaaataaaaatataaactcatcaaacggaaatacatgagtctttttcacatcttaaat 169
Db    1067 ATTWCWCAAMTAATMAATCTAAAAATTYATMWAAACAAAACAAATAAAMAAAAAAAAAAT 1008

Qy     170 cgaataatagtaattttgtagaatatggatcttattgcggagatgggtgtaactctagaa 229
Db    1007 AAAAAAAAAAAAAMTCWHAAATTCAAAAACAATNAMACTCTTATTACCWATCTACTAAAA 948

Qy     230 ttacaaaaaatataataaaatataaataaaaatgatagaaaatcaccaagacaaatttttag 289
Db     947 TWAYAAAAAAAAAAAAAAAAAAAAYATCTATATAAAAAACATATACCCAWAAAAAAAAAAA 888

Qy     290 aggagtataaaaaaaggaaacaaggtattatagcaggatactatggttcatggaacagtc 349
Db     887 AAYTTAAATAAATAAAAAAAAAAAAAAAAATAAAATAAATNATTTYTTAAATTAAAWAHTT 828

Qy     350 aaggtgatagagcaaaacatatgattgattcaaacccaatggtgtcaattttatatattg 409
Db     827 ATTTWWAATAAYTTTTHAATATAATTTATAATYATAACTTATTYWAATTAAAATNTTYTT 768

Qy     410 catttgctcgcattaatatgttatatgatgtatctagaccatttaatggaagacaaagat 469
Db     767 AATTTTWTAWATAATAWTAATTAAAAAAAATTTTTAAAAAATATAATTTAANAAATATTA 708

Qy     470 tcctattaagaaaacacggcttagaatatgaaacctatggtatgatgcttaatgaaatta 529
Db     707 TATTTTAAATATTTAATTTTTTTAATAATTTTAAAATATTTTTTTTYCTATTAAAAAATA 648

Qy     530 gacgtatcagaaaagtacgtccagatgtaattattcttttatccttaggtggagaaacct 589
Db     647 TAAACAATAAAAATWAAATTATTAWAACAANAACBATATTATATTAWATTTTATTAKANT 588

Qy     590 atatgatagatatagaaaaagaaattgattatgtggataaaatattgaagcttgtt---a 646
Db     587 ATTAATTTATTTTAAAAAATTTTTTKKKKKKKTBSCACCAACYAAAAANTTTMYTYCCCC 528

Qy     647 atgattttgatttagatggtgtagatattgactgggaaccacatgggaagttttacaact 706
Db     527 CCCCCTTTTTTTWTTTTTTTTNTNANNANNATAANATATTTTWTAAAAATWTATTTTTTW 468
```

```
Qy     707 taaatgaattaaatttttcaaattattatattaaattaattaacttgttaagaaaaacta 766
           :||:| | :| | ||| | || |: : | |||  |:: | ::|     ||| | |   ||
Db     467 WAAWTAACYTKATTTTATTAATTWTWAAAATTYYAKWTAAWDAAAAAATAAAATATTTTA 408

Qy     767 ttccggaagaaaagttaatttcaattttct---ggttcatcaaatgctgcattatcatgcg 823
             : | : : :::: :|: |:|||:|   :::|  |   |         |||| :::
Db     407 AAAAKTATBTDTDKKKKKKTWWADTTTBTYABSKKTTTTNNTANNNATGTATATCTASSS 348

Qy     824 tttcaggagttgcatctttctgtaaagatgaagaatctccatataacactaaatttttgt 883
               |  ||| :| | :    ||     | |||  |  :| ||    : |||||| :
Db     347 SSSKKGANATTGTWTNTCBTGTAAATNBNWATGAANNTATWTTTATTTTWTAATTTTTTY 288

Qy     884 ctgaacaaatagaaacaaataaagaattacatagggcagcagcgatgttatcagcaggaa 943
           |      |          |       |          ::|  |   || |      |  :
Db     287 TTYTTTKKBTNTSBBTSSCTNNNNNTNTCTSSSCCSSATAATATATATATATCCCCCSCB 228

Qy     944 cttttattaatatttttaatacagcaaaggagaaaatagatcttgtatttattcaaacat 1003
           :| |   | ||  ||  | ||    |||    || ||   |:|| |||||| |
Db     227 MTATMTATTATTATTAAATTATTAAAAAATTTAACATGTBTMTTATATTTACTACTTATA 168

Qy    1004 acaatttagaaactacaaatccagatataatggtagatatgtacttatcccatttatatt 1063
           |  ||  |  || : :      | ||:|  |   |  |||||  |  | |  |   ||
Db     167 AATATAAAATAAAHTHTYTATTATATWTTTTWTAAATTATGTATATTCCGCGATATCATC 108

Qy    1064 ttggtttaaaat 1075
             |    | | ||
Db     107 ATATGATGATAT 96
```

```
RESULT  12
CNS0006J
LOCUS       CNS0006J     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01M22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL062049
VERSION     AL062049.1  GI:4938511
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR01M22"
                      /note="end : TET3"
BASE COUNT      333 a    162 c    148 g    177 t    281 others
ORIGIN
```

```
  Query Match             4.4%;  Score 78.4;  DB 13;  Length 1101;
  Best Local Similarity   30.2%;  Pred. No. 0.014;
  Matches  191;  Conservative  158;  Mismatches  276;  Indels    7;  Gaps    2;

Qy      70 aaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaataaaact 129
           |||  | ||||: | | ||| |:   | || :|||:: |  |:  |:|  ||| ||||
Db     472 AAAAAATAAAADGAAAAAAAAAARARRAATGARAATRWRARAGRTTAWTTGAAAAAAAAAA 531

Qy     130 catcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgta 189
            |:   |  |||| | |  |    | |: |  | || ::  ::|  ||   :   :   |
Db     532 AAWTTTAAAGAAAAAAA--AAAWATATKAAATTTTTTRRRTRRAGTATTTRDTRAKAAAA 589

Qy     190 gaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaatataaaaat 249
           || |   ||   |    | | |      : || | |   |  | |||||| || ||:|
Db     590 AAAAAAATATTAGAAAARGAAAAAAAAAARAAAGTGTTTGAAGAAAAAAAAAAAAAADATA 649

Qy     250 ataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaaggaaa 309
           |  |:  |||| |  | ||||| |  ||:| ::  :|   |  |  | |:|||:||| |
Db     650 AAKAWAKAAAAGGGAAAAAAATAAAAAAKAAWRTGKTATAAAAAAAAAAWAAARAGGGAK 709

Qy     310 caaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacat 369
             |: |  : :| :||  |  |   :| ::::    :: |||:  : | || |||| |
Db     710 AGGGKRAGARARGGRGGGRAWRAAAAKTKRKKKTKTKRKRAAGRRARRAWAGAAAAAAAR 769

Qy     370 atgattgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatg 429
           |  :    |   :|   ::|  ::: :  |: | |:|  : :::|  : :  |::  |
Db     770 AAADAATKTGKTWAKAAWRTAAWKKDWKATDAAAAKAAARTRDWTTAKAKDKTKWGAAAA 829

Qy     430 ttatatgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggc 489
           | |:  || : :   ::    |  :: :|| :  :|  |:  .  |: |  |:::  |
Db     830 TGAWGAGAARGRRWGRKGKDGTRARRARGAGDGWDAKAAWAWAAAWAATTARWDTATAAD 889

Qy     490 ttagaatatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgt 549
           : :  |::::||| :::::: :|   | :: :|:| |::|  ::  |:| |  ::  :
Db     890 DRRAKAWDKRAAAAADRDKDKGRRAGWGTGWRRRARARTRKAKAKRRGARAGAKAKRRAK 949

Qy     550 ccagatgtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaaa 609
             | :|:    :: ::::|   :|:|||::: :|    | ::| |: :::: ||::
Db     950 GDAWKDGW-----KKKWDKDAAWAAKAKGTGDWKKRARKAKTGWKARADKDRWRAAADDT 1004

Qy     610 gaaattgattatgtggataaaatattgaagcttgttaatgattttgatttagatggtgta 669
           ::: :  ::|:::: || : :| ::   |  ::|| :|::  : :|::: ||:    :
Db    1005 RWRTKAWWWTDWRWATATKDKDTKDKAKKGKAWKTTTRTDRAAKAKAKDDTAATRTATAW 1064

Qy     670 gatattgactgggaaccacatgggaagtttta 701
           :: :::    : :|   | |: ::||::: :|
Db    1065 KDADDKAKARDAKAKWGADAKADRAADDWRDA 1096
```

```
RESULT  13
CNS01U9O/c
LOCUS       CNS01U9O     524 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            196C24 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL167541
VERSION     AL167541.1  GI:7805598
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 524)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 524)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="196C24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG196BB12SP1~end :
                     PUC-Ori"
BASE COUNT      124 a     15 c     24 g    298 t     63 others
ORIGIN


  Query Match             4.4%;  Score 77.6;  DB 13;  Length 524;
  Best Local Similarity   42.7%;  Pred. No. 0.024;
  Matches  165;  Conservative   47;  Mismatches  172;  Indels    2;  Gaps    1;

Qy      70 aaaggaaaaaataatataaataattcattgggaataatacgggaaaataaaaataaaact 129
           |||  |||||| || | ||| ||   |      ||||| |    |||| ||| ||||||
Db     457 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAATAAAAAAAATATAAAAAA 398

Qy     130 catcaaacggaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgta 189
           |  | |   || :| :| |    |: : | |  |||||   |:||| |  ||  :  |
Db     397 AAAAATAAAWAATWATWTWATAAATWAWAAAAAATTAAAAWWAWTAAAAWAAAAAWATTT 338

Qy     190 gaatatggatcttattgcggagatgggtgtaactctagaattacaaaaaataataaaaat 249
           ||: |  :| ||: |      :|   : | | | || || :| ||: |: || ||:||
Db     337 TAAWTT--WTATTWATATTTTAWTTTTWTTTATTTTAAAAAWATAAWTAWAAAAAARAAA 280

Qy     250 ataaataaaaatgatagaaaatcaccaagacaaatttttagaggagtataaaaaaaggaaa 309
           | ||| |||||  | | ||||  |  || | |||    | |  |  |:|:|||||
Db     279 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAATTAAAAWAWAAAAAAWTTT 220

Qy     310 caaggtattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacat 369
           |   |||| |      ||  ||:  |  ||       | ||    |:| |  |||| |
Db     219 TATATTATTTTTTATATATTTTAWAWTATATTTTTATTTTAAAAAAAWAAAAAAAAAAAAA 160

Qy     370 atgattgattcaaacccaatggtgtcaatttttatatattgcatttgctcgcattaatatg 429
           |  |:| |:  |||   ||   | |   ||:::| |::|  ::::  :   :::::::
Db     159 AAAAWTAAWATAAAAAAAAAAATTTTTTTTWWWTTTWWTTTWWWWWWWWWWWWWWWWST 100

Qy     430 ttatatgatgtatctagaccatttaa 455
           || |:: :  :::  | |: :  |:|
Db      99 TTTTWWWWAAWHWAAAAAMAMAATWA 74


RESULT  14
CNS009G1
LOCUS       CNS009G1      876 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19J14 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053529
VERSION     AL053529.1  GI:4935018
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .876
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19J14"
                     /note="end : TET3"
BASE COUNT      335 a     54 c     57 g    325 t    105 others
ORIGIN


  Query Match             4.4%;  Score 77.4;  DB 13;  Length 876;
  Best Local Similarity   40.3%;  Pred. No. 0.022;
  Matches  210;  Conservative   43;  Mismatches  265;  Indels    3;  Gaps    1;

Qy     211 gatgggtgtaactctagaattacaaaaaataataaaaatataaataaaaatgatagaaaa 270
           :: :|  | ||    | ||    ||| || |  |||| | ||| |||||       | |
Db     358 KVADGCRGCAAAAANAAAAAACAAAANAAAANAAAAAAAAAAAAAAAAAANANANANANA 417

Qy     271 tcaccaagacaaattttagaggagtataaaaaaaggaaacaaggtattatagcaggatac 330
           |  || | :||   |  | |  |  || | |  : |     ::|||||  |  | |
Db     418 AAAAAAATATWAAAAATTAANAANAAATAATATATTWTATTNAAWWTTATATAAATAAAT 477

Qy     331 tatggttcatggaacagtcaaggtgatagagcaaaacatatgattgattcaaacccaatg 390
           |     | |:  |:  |  |  | ||| :   || |    | | |   |||   ||
Db     478 TNATWATWAWTTAWATATATATATWATAAWTTTTAAAAATATAATAAAATAAAAWWAAAA 537

Qy     391 gtgtcaattttatatattgcatttgctcgcattaatatgttatatgatgtatctagacca 450
           || | ||||:|    |||   :   ||   ||| : :|:  |: ||: || :  |
Db     538 TAWAAAAATATATAWAAAAAATTAWAWATAATATTTATAWAWTWAAAWTTAWWTAAWTTA 597

Qy     451 tttaatggaagacaaagattcctattaagaaaacacggcttagaatatgaaacctatggt 510
           | |||:     | | | |     : ||| ||   |     |  |   : ::|  |||  |
Db     598 TATAAWAATTTATATATAAAWAAWATAAAAATTAATANAATTTATATWAWWATATATAWT 657

Qy     511 atgatgcttaatgaaattagacgtatcagaaaagtacgtccagatgtaattattctttta 570
           ||  :    :|| | |||  :  ||| :  :|| ||     | || ||  |||| || ||
Db     658 ATATWTTANWATAATATTTTWAATATWWTTWAAATATTNWTATATATATATATTTTTATA 717

Qy     571 tccttaggtggagaaacctatatgatagatatagaaaaagaaattgattatgtggataaa 630
           : |  |  |  |   :||||     ||| : |:| | :|:    |||: :  :| |
Db     718 NAAWAATATATAWTANATWATATATATTATANWTAWATANWAWAANTTTAWTWAAWTTAT 777

Qy     631 atattgaagcttgttaatgattttgatttagatggtgtagatattgactgggaaccacat 690
           :|::  |   ||  | || ||| | || || ||    || ||| |   |   |   | ||
Db     778 WTWWNTA---TTAATTATWATTATAATATATATATANTATATAATATATATWATWTATAT 834

Qy     691 gggaagttttacaacttaaatgaattaaatttttcaaatta 731
           |  | |:| |  ||  || |  || ::| |   |  ||
Db     835 ATTATATATWATATTTTTTATTANATATWWTNTATTATATA 875
```

```
RESULT  15
BF259858/c
LOCUS       BF259858      839 bp    mRNA            EST       23-FEB-2001
DEFINITION  HVSMEf0020H09f Hordeum vulgare seedling root EST library HVcDNA0007
            (etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMEf0020H09f, mRNA sequence.
ACCESSION   BF259858
VERSION     BF259858.2  GI:13120301
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 839)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     On Nov 16, 2000 this sequence version replaced gi:11188971.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 825.
FEATURES             Location/Qualifiers
     source          1. .839
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEf0020H09f"
                     /clone_lib="Hordeum vulgare seedling root EST library
                     HVcDNA0007 (etiolated and unstressed)"
                     /tissue_type="Seedling root"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; For
                     more details on library preparation and sequence analysis
                     see http://www.genome.clemson.edu/projects/barley/ To
                     order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      275 a     46 c     56 g    451 t     11 others
ORIGIN
```

Query Match 4.4%; Score 77.2; DB 11; Length 839;
Best Local Similarity 46.7%; Pred. No. 0.024;
Matches 263; Conservative 0; Mismatches 297; Indels 3; Gaps 1;

```
Qy     75 aaaaaataatataaataattcattgggaataatacgggaaaataaaaataaaactcatca 134
          |||| | || |||||  ||| |||    |||| |    |||| ||||| ||||  ||  |
Db    787 AAAATAAAAAATAAAATATTTATTATATATAAAAAATAAAAAAAAAAAAAAAAAACAAAA 728

Qy    135 aacggaaatacatgagtctttttcacatcttaaatcgaataatagtaattttgtagaata 194
          ||   ||  | || |          | |    ||||   ||||  ||    |  || |||
Db    727 AAAAAAAGAAAATAAAAAAAAAAAAAAAAAAAAAAAATAATATAAATGTTTAATATTATAATT 668

Qy    195 tggatcttattgcggagatgggtgtaactctagaattacaaaaaatataataaaaatataaa 254
          |   |  | |     |        |||    | ||||| |||||| || |||||   |||
Db    667 TAATTTATTTAAATAATTATTAATTAAAATAAAAATTAAAAAAAAAAAAAAAAA---AAA 611

Qy    255 taaaaatgatagaaaatcaccaagacaaattttagaggagtataaaaaaaggaaacaagg 314
           ||||   | | ||||  |  || | |||    | |  |  | |||||||  ||| ||
Db    610 AAAAATAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 551

Qy    315 tattatagcaggatactatggttcatggaacagtcaaggtgatagagcaaaacatatgat 374
          ||  | |  |  | |  |    |  |  || |   ||    ||| |  |||| | |  |

Db    550 TAAAAAATAAAAAAAAAAAAAATATTAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAA 491

Qy    375 tgattcaaacccaatggtgtcaattttatatattgcatttgctcgcattaatatgttata 434
          |   |||    |       ||    |  ||     |||  |   ||    || |  |
Db    490 AAAAAAAAAAAATAAAAAAAAAAAAAAAAAATTAAAAATTTTTTTAATATATTAATATNNTT 431

Qy    435 tgatgtatctagaccatttaatggaagacaaagattcctattaagaaaacacggcttaga 494
          |  | | |    | |||||     | |      ||  | |||   ||      |  |
Db    430 TTNTNTTTTATAATTATTTATATNTAAATTTTTTTTATTTTTATATAATTTTTTTTATAA 371

Qy    495 atatgaaacctatggtatgatgcttaatgaaattagacgtatcagaaaagtacgtccaga 554
          |  | |||  | |  ||  |   |      ||||     | |   ||   |   |
Db    370 AATTTAAAATTTTTATANTAAATTNNTAATAATTTATAATTTTTTAATTTTTATTTTTAT 311

Qy    555 tgtaattattcttttatccttaggtggagaaacctatatgatagatatagaaaaagaaat 614
          | || |  ||   |||| | |   | |||||   | |||| | | |||  || |  | |
Db    310 TTTATTATTTTAGTTATATTAATAAGTAGAAATAAAGATGAAAAANATATTAATAATATT 251

Qy    615 tgattatgtggataaaatattga 637
          |  | ||    | |||| ||| |
Db    250 TAGTGATTAATAAAAAAAATTAA 228
```

Search completed: March 21, 2002, 14:56:13
Job time: 1904 sec

OM of: US-09-579-383-2 to: A_Geneseq_1101:* out_format : pfs

Date: Mar 21, 2002 4:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09579383/runat_20032002_151238_29288/app_query.fasta_1.1849
-DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09579383_@CGN1_1_87 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-579-383-2
Query length: 1764
Database: A_Geneseq_1101:*
Database sequences: 522463
Database length: 74073290
Search time (sec): 104.710000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49986	+	3130.00	5266.42	3.9e-286	587	
/SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49985	+	635.00	1053.49	2.8e-51	378	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY55791	+	148.00	217.88	3.0e-05	1215	
/SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB84784	+	148.00	217.88	3.0e-05	1215	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18255	+	144.00	206.85	8.3e-05	1817	
/SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR64823	+	142.50	216.77	7.5e-05	561	
/SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:AAW09402	+	142.50	216.77	7.5e-05	561	
/SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW29178	+	142.50	216.77	7.5e-05	561	
/SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW60179	+	142.50	216.77	7.5e-05	561	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18180	+	139.50	196.76	0.0002	2295	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18253	+	138.00	188.41	0.0004	3973	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18272	+	135.50	189.09	0.0006	2500	
/SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:AAW02159	+	130.00	191.04	0.0013	866	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52307	+	130.00	191.04	0.0013	866	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52533	+	126.50	190.76	0.0023	509	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB24129	+	126.50	184.32	0.0029	934	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18171	+	126.50	176.36	0.0038	1979	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18172	+	126.00	173.10	0.0046	2485	
/SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:AAR48981	+	125.50	188.04	0.0030	561	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18275	+	125.00	175.35	0.0050	1714	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18311	+	123.00	179.95	0.0059	807	
/SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY31946	+	122.50	161.78	0.0118	4134	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52535	+	121.00	181.86	0.0076	490	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18301	+	121.00	167.97	0.0121	1817	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18173	+	120.00	179.81	0.0095	507	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18254	+	120.00	164.22	0.0161	2206	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18287	+	119.50	164.43	0.0174	1997	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18183	+	119.00	169.09	0.0160	1188	
/SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW30763	+	118.50	168.34	0.0178	1178	
/SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:AAR55694	+	118.00	159.98	0.0256	2391	
/SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY33683	+	117.50	166.41	0.0223	1205	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90882	+	117.50	166.41	0.0223	1205	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18315	+	117.00	158.34	0.0318	2380	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18222	+	116.50	156.99	0.0360	2496	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52023	+	116.00	160.64	0.0345	1634	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY51652	+	116.00	160.64	0.0345	1634	
/SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:AAR07503	+	115.00	161.76	0.0389	1254	
/SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW24575	+	115.00	161.76	0.0389	1254	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY54040	+	114.50	171.30	0.0305	471	
/SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR95011	+	114.50	154.18	0.0546	2366	
/SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW68388	+	114.50	154.18	0.0546	2366	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18168	+	114.00	169.04	0.0357	538	
/SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18244	+	113.50	159.30	0.0538	1245	
/SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:AAW98149	+	113.50	152.86	0.0669	2285	
/SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW44842	+	112.50	163.55	0.0546	711	

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49986

seq_documentation_block:
ID AAB49986 standard; Protein; 587 AA.
XX
AC AAB49986;
XX
DT 13-MAR-2001 (first entry)
XX
DE P. gallinaceum chitinase SEQ ID NO: 4.
XX
KW Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KW veterinary disease; arthropod pest.
XX
OS Plasmodium gallinaceum.
XX
PN WO200073488-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14536.
XX
PR 28-MAY-1999; 99US-0136508.
PR 03-FEB-2000; 2000US-0180051.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Vinetz JM;
XX
DR WPI; 2001-061553/07.
DR N-PSDB; AAC89669.
XX
PT New nucleic acid encoding a Plasmodium species chitinase is useful for
PT preventing transmission of malaria by mosquito feeding on subject that
PT may harbor Plasmodium species organisms -
XX
PS Claim 29; Page 125-127; 137pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC organisms are the cause of malaria in humans. The sequences are useful
CC in the prevention and treatment of malaria, fungal diseases, parasitic
CC diseases and veterinary diseases, in preventing the transmission of
CC malaria and in the control of arthropod pests in agriculture.
XX
SQ Sequence 587 AA;

alignment_scores:
Quality: 3130.00 Length: 587
Ratio: 5.332 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-579-383-2 x AAB49986 ..

Align seg 1/1 to: AAB49986 from: 1 to: 587

```
      1 ATGAATTTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTC 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetAsnPheLysIleSerIlePheLeuIleIleValSerIleLeuTyrSe 17

     51 TGCAAATTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTCATTGG 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     17 rAlaAsnSerArgThrLeuLysGlyLysAsnAsnIleAsnAsnSerLeuG 34

    101 GAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATACATGAG 150
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  34 lyIleIleArgGluAsnLysAsnLysThrHisGlnThrGluIleHisGlu 50

 151 TCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATC 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 SerPheSerHisLeuLysSerAsnAsnSerAsnPheValGluTyrGlySe 67

 201 TTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATA 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 rTyrCysGlyAspGlyCysAsnSerArgIleThrLysAsnAsnLysAsnI 84

 251 TAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAA 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 leAsnLysAsnAspArgLysSerProArgGlnIleLeuGluGluTyrLys 100

 301 AAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCA 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 LysArgLysGlnGlyIleIleAlaGlyTyrTyrGlySerTrpAsnSerGl 117

 351 AGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTT 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 nGlyAspArgAlaLysHisMetIleAspSerAsnProMetValSerIleL 134

 401 TATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCA 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 euTyrIleAlaPheAlaArgIleAsnMetLeuTyrAspValSerArgPro 150

 451 TTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGA 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 PheAsnGlyArgGlnArgPheLeuLeuArgLysHisGlyLeuGluTyrGl 167

 501 AACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTC 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 uThrTyrGlyMetMetLeuAsnGluIleArgArgIleArgLysValArgP 184

 551 CAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGAT 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 roAspValIleIleLeuLeuSerLeuGlyGlyGluThrTyrMetIleAsp 200

 601 ATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGA 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 IleGluLysGluIleAspTyrValAspLysIleLeuLysLeuValAsnAs 217

 651 TTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTT 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 pPheAspLeuAspGlyValAspIleAspTrpGluProHisGlyLysPheT 234

 701 ACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAAC 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 yrAsnLeuAsnGluLeuAsnPheSerAsnTyrTyrIleLysLeuIleAsn 250

 751 TTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTC 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuLeuArgLysThrIleProGluGluLysLeuIleSerIleSerGlySe 267

 801 ATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAG 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 rSerAsnAlaAlaLeuSerCysValSerGlyValAlaSerPheCysLysA 284

 851 ATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 spGluGluSerProTyrAsnThrLysPheLeuSerGluGlnIleGluThr 300

 901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTAT 950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 AsnLysGluLeuHisArgAlaAlaAlaMetLeuSerAlaGlyThrPheIl 317

 951 TAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 eAsnIlePheAsnThrAlaLysGluLysIleAspLeuValPheIleGlnT 334

1001 CATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 334 hrTyrAsnLeuGluThrThrAsnProAspIleMetValAspMetTyrLeu 350

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 SerHisLeuTyrPheGlyLeuLysTyrAsnIleThrIleIleLeuGlyPh 367

1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTAT 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 eSerLeuGluHisAsnArgGlyGlyPheSerProGluAsnLysGluLeuL 384

1151 TAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAATAGG 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 euGluLeuValGlyLysThrIleHisAspLysAsnGlnAsnAsnAsnArg 400

1201 GCAGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAAC 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 AlaAspGlyIleGlyIleTrpHisLeuPheMetLysGluGlnLeuProTh 417

1251 TGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGGAAACATTTAA 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 417 rGlySerPheAspValAspIlePheLeuThrAsnIleTrpLysHisLeuA 434

1301 ATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAA 1350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 434 snProGluValGlnThrProLysAspLeuThrIleThrGluAsnProGlu 450

1351 GACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCAT 1400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 451 AspCysSerThrIleAspGluTyrValProGlyLeuValIleProThrIl 467

1401 AGGGATATATTACAAACACAATGATGCTATATGGAAAACTAGATCTTATT 1450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 467 eGlyIleTyrTyrLysHisAsnAspAlaIleTrpLysThrArgSerTyrS 484

1451 CAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAAAGTA 1500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 484 erIleHisAlaProGlyValAspArgTyrGluTrpAspLeuValLysVal 500

1501 TGCTATGAAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGA 1550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 501 CysTyrGluLysIleCysAspGlyLysAlaAlaHisTyrTyrAsnThrAs 517

1551 CTATAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTA 1600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 517 pTyrLysGluSerSerIleIleIleTrpLysGlyGluProTyrLeuIleL 534

1601 AATGGTGGCAACAAGGACCTCCGGAAGGTCAGGCACTAGAGTCATACACA 1650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 534 ysTrpTrpGlnGlnGlyProProGluGlyGlnAlaLeuGluSerTyrThr 550

1651 AAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATGGAATAAAAAATA 1700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 551 LysLeuAspAlaSerLysCysProGlyIleGluGluTrpAsnLysLysTy 567

1701 TCCACATAAACCACTAGAAGTAGAGGAACAATATGAACAAGAAGTGGATT 1750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 567 rProHisLysProLeuGluValGluGluGlnTyrGluGlnGluValAspL 584

1751 TACCATTACAA 1761
     ||||||||||||
 584 euProLeuGln 587
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49985

seq_documentation_block:
ID AAB49985 standard; Protein; 378 AA.
XX
AC AAB49985;
XX
DT 13-MAR-2001 (first entry)
XX
DE P. falciparum chitinase SEQ ID NO: 3.
XX

```
KW   Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KW   veterinary disease; arthropod pest.
XX
OS   Plasmodium falciparum.
XX
PN   WO200073488-A1.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14536.
XX
PR   28-MAY-1999;   99US-0136508.
PR   03-FEB-2000; 2000US-0180051.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Vinetz JM;
XX
DR   WPI; 2001-061553/07.
DR   N-PSDB; AAC89668.
XX
PT   New nucleic acid encoding a Plasmodium species chitinase is useful for
PT   preventing transmission of malaria by mosquito feeding on subject that
PT   may harbor Plasmodium species organisms
XX
PS   Claim 29; Page 124-125; 137pp; English.
XX
CC   The present invention provides the protein and coding sequences of the
CC   Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC   organisms are the cause of malaria in humans. The sequences are useful
CC   in the prevention and treatment of malaria, fungal diseases, parasitic
CC   diseases and veterinary diseases, in preventing the transmission of
CC   malaria and in the control of arthropod pests in agriculture.
XX
SQ   Sequence   378 AA;


alignment_scores:
            Quality:  635.00              Length:    345
              Ratio:  2.560                Gaps:      5
 Percent Similarity:  71.884    Percent Identity: 37.391

alignment_block:
 US-09-579-383-2 x AAB49985   ..

 Align seg 1/1  to: AAB49985  from: 1  to: 378

     265 AGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGG 314
         ||||||:::|||||||:::|||:::::   :::|||:::      :::||
      37 ArgLysAsnProArgGluIleIleLysThrPheLysGluSerGlyLysGl 53

     315 TATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAA 364
         |||||||   |||||||||   ||||||   |||   :::   :::   |
      53 yIleIleGlnGlyTyrTyrProSerTrpValSerTyrAsnHisAsnLeuL 70

     365 AACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTT 414
         ||         |||   ||||||   ::::::::::::::::::::|||
      70 ys.........AspLeuAsnProAsnLeuAsnValValHisMetSerPhe 83

     415 GCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGAAGACA 464
         |||::::::::::::   ||||||      :::   :::   |||
      84 AlaLysMetAspLeuSerTyrAspSerIleGluSerIleValGlySerPr 100

     465 AAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGA 514
             ::::::            ||||||||||||         :::
     100 oLeuLeuPheLysSerLeuIleGlyLeuGluTyrIleGlyLeuAsnGluT 117

     515 TGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATT 564
           :::|||:::      ::::::||||||:::||||||||||:::|||:::
     117 yrPheAsnAspAlaMetAsnLeuArgLysAlaArgProAspIleIleMet 133

     565 CTTTTATCCTTAGGTGGAGAAACCTAT...ATGATAGATATAGAAAAAGA 611
         ||||||||||||||||||||||||||         ::::::::::::
```

```
     134 LeuLeuSerLeuGlyGlyGluThrTyrHisProSerSerPheAspSerAl 150

     612 AATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAG 661
           ::::::   |||:::||||||   :::||||||:::::::::   :::|
     150 aLeuAsnAlaValGluLysIleAlaAsnLeuValAspGluLeuGlyPheA 167

     662 ATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAAT 711
         |||||:::|||:::|||:::||||||:::|||:::|||   :::||||||
     167 spGlyIleAspValAspTyrGluProAsnGlySerPheAspGlyLeuAsn 183

     712 GAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAA 761
         :::   :::   ::::::::::::::::::   ::::::   ||||||::
     184 AspLysGluLysAlaAspPhePheValGlnTyrValThrLysLeuArgGl 200

     762 AACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTG 811
         :   :::   ::::::||||||||||||||||||   ||||||||||:::|
     200 uTyrMetCysAspAspLysLeuIleSerIleSerGlnSerSerAsnGlyA 217

     812 CATTATCATGCGTT......TCAGGAGTTGCATCTTTCTGTAAAGATGAA 855
         ||||||||||:::       :::         :::::|||   |||:::
     217 laLeuSerCysIleGlyPheAsnAspProLysLysIleCysMetAspAsp 233

     856 GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAA 905
         |||:::|||||||||:::|||:::::::::::::       ::::::::::||
     234 GluAlaProTyrAsnSerLysTyrPheAsnLysPro...AspValLysLy 249

     906 AGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATA 955
         |||||||   |||||||||   |||   |||||||||       |||   :
     249 sGluLeuLeuArgAlaAlaGlnMetAlaSerAlaGlyGlyAlaIleTyrL 266

     956 TTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATAC 1005
         :::::|||:::   |||:::   ||||||:::||||||:::|||||||:::
     266 euMetAsnAsnLeuLysAspMetIleAspMetValPheValGlnThrPhe 282

    1006 AAT...TTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTATC 1052
         |||      ::::::|||:::      :::|||   :::::::|||   ||
     283 AsnTyrThrAsnSerThrAspSerThrValMetLysGluLeuTyrAspSe 299

    1053 CCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTT 1102
         |:::   |||:::|||   ||||||:::   :::||||||:::||||||:
     299 rTyrAlaTyrTyrGlyLysLysTyrAspTyrValIleIleMetGlyPheT 316

    1103 CATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTA 1152
         ::|||                |||:::|||::::::|||   |||:::
     316 hrLeuMetPheProSerThrProPheAsnProAsnAspLysMetLeuVal 332

    1153 GAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAATAGGGC 1202
         :::   :::|||      :::      :::|||:::   |||:::||||
     333 LysSerIleGlyAspPheValLysThrGluAsnLysLeuAsnLysArgAl 349

    1203 AGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAACTG 1252
         |||||||:::|||:::|||   |||         :::         :::
     349 aAspGlyPheGlyLeuTrpSerLeuSerSerAspAsnAlaAlaHisAsnG 366

    1253 GATCATTCGATGTAGATATTTTTCTTACAAATATT 1287
           ::::::   ::::::   |||:::   ::::::
     366 luGlnLeuAlaIleGluTyrPheValGluSerLeu 377

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY55791

seq_documentation_block:
ID   AAY55791 standard; protein; 1215 AA.
XX
AC   AAY55791;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Superheat-resistant chitinase.
XX
KW   Superheat-resistant; thermostable; chitinase; hydrolysis; chitin;
KW   chitobiose.
XX
OS   Pyrococcus sp.
```

```
XX
PN   JP11313688-A.
XX
PD   16-NOV-1999.
XX
PF   19-FEB-1999;   99JP-0041547.
XX
PR   20-FEB-1998;   98JP-0039285.
XX
PA   (BEAB-) BE ABLE KK.
XX
DR   WPI; 2000-057358/05.
DR   N-PSDB; AAZ39542.
XX
PT   A superheat-resistant chitinase -
XX
PS   Claim 1; Page 7-12; 19pp; Japanese.
XX
CC   This represents a superheat-resistant chitinase. The protein can be
CC   expressed by standard recombinant methodology. Chitinase hydrolyzes
CC   chitin, and decomposes chitin to form chitobiose.
XX
SQ   Sequence   1215 AA;


alignment_scores:
            Quality:  148.00              Length:    647
              Ratio:  0.576                Gaps:     28
 Percent Similarity:  39.722   Percent Identity: 17.311

alignment_block:
 US-09-579-383-2 x AAY55791   ..

 Align seg 1/1  to: AAY55791  from: 1  to: 1215

     160 CATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGG 209
         |||||||||:::|||:::::   :::                    ::
     139 HisLeuLysAlaAsnThrThrTyrTyrTyrGlyValValProValLeuAl 155

     210 AGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAA 259
         :||||||
     155 aAspGly..................................... 157

     260 ATGATAGAAAATCACCAAGACAAATT.................TTAGAG 291
           :::|||   ||||||   ::::::                |||||
     158 ..SerArgGlySerProSerAsnValLeuAlaIleThrThrProLeuGlu 173

     292 GAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATG 341
           |||:::                :::      ||||||   |||||
     174 ProTyrArg.................ValIleValTyrTyrIleSerTr 184

     342 GAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGG 391
         |:::      :::         |||      :::::::::   |||
     184 pGlyArgTyrAla......ArgLysPheTyrValSerAspIleProTrpG 199

     392 TGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTA 441
           ::::::   ::::::   :::|||   :::::::::      |||
     199 luLysValThrHisValAsnTyrAlaPheLeuAspLeuLysGluAspGly 215

     442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
         :::      |||
     216 ThrValAlaPhe..................................... 219

     492 AGAATATGAAACCTATGGTATGATGCTTAAT......GAAATTAGACGTA 535
             |||:::|||||||:::      ||||||         :::::::::
     220 ....TyrAspThrTyrAlaAspProLeuAsnLeuGluAlaMetLysGluT 235

     536 TCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAA 585
           ::::::      |||   |||   :::|||:::|||:::||||||
     235 yrLysArgLysTyrProAlaValLysValLeuIleSerValGlyGlyTrp 251

     586 ACC.....................TATATGATAGATATAGAAAAAGAAAT 614
         |||                        |||      |||:::

     252 ThrLeuSerLysTyrPheSerValValAlaAlaAspProAlaLysArgGl 268

     615 TGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATG 664
           :::::::::      :::::::::::::::   ::::::|||||||
     268 nArgPheAlaGluThrAlaIleGluIleLeuArgLysTyrAsnLeuAspG 285

     665 GTGTAGATATTGACTGGGAA.........CCACATGGGAAGTTTTACAAC 705
         ||:::|||||||||||||||              |||         |||
     285 lyIleAspIleAspTrpGluTyrProGlyGlyGlyGlyMetAlaGlyAsn 301

     706 TTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTT 755
           ::::::   :::   ::::::   ::::::   |||:::::   ||
     302 TyrGluSerProAspAspGlyLysAsnPheValLeuLeuLeuLysAspLe 318

     756 AAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAA 805
         ||||::::::::::   :::
     318 uArgGluAlaLeuAspLysAla.......................... 325

     806 ATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAA 855
                                                :::|||::::::
     326 ....................................AlaLysGluAsp 329

     856 GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAA 905
         ::::::   |||
     330 HisLysAspTyr..................................... 333

     906 AGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATA 955
                                 :::|||:::|||:::|||      :::
     334 .....................LeuLeuThrAlaAlaThrProAlaAspP 343

     956 TTTTTAATACA.....................GCAAAGGAGAAAATAGAT 984
               ::::::                    |||::::::   :::|||
     343 roValLysAlaGlyArgIleAspTrpValGluAlaSerLysTyrLeuAsp 359

     985 CTTGTATTTATTCAAACATAC............................. 1005
           :::   |||:::|||||||
     360 SerIleAsnIleMetThrTyrAspTyrHisGlyAlaTrpGluThrIleTh 376

    1006 ...............................AATTTAGAAACTACAAATC 1024
                                        |||         |||:::
     376 rGlyHisLeuAlaProLeuTyrCysAspProAsnAlaProTyrThrAspG 393

    1025 CAGATATA........................ATGGTAGATATGTACTTA 1050
           ::::::                       |||:::   |||:::
     393 luAsnValLysTyrHisPheCysValAsnTyrThrValGlnTrpTyrIle 409

    1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
         :::|||:::           |||   :::|||||||:::   |||   ||
     410 GlnHisVal......ProAspLysThrLysIleThrValGlyLeuProPh 424

    1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTAT 1150
         |      :::::::::   ::::::      |||||||||:::   |||
     424 e...TyrSerArgSerPheAlaAsnValProProGluAsnAsnGlyLeuT 440

    1151 TAGAATTGGTAGGAAAAACA............................. 1170
           :::      :::   |||
     440 yrGlnProPheSerGlyThrProAlaGlyThrTrpGlyProAlaTyrGlu 456

    1171 ..........................ATACATGATAAAAATCAAAATAA 1193
                                    :::   :::|||||||||:::::
     457 ThrTyrGlyValMetAspTyrTrpAspValAlaGluLysAsnGlnSerSe 473

    1194 T.............................................. 1194
         :
     473 rGluTyrGluTyrHisTrpAspProIleAlaGlnValAlaTrpLeuTyrS 490

    1194 .............................................. 1194

     490 erProSerLysArgIlePheIleThrPheAspAspProArgAlaIleGly 506

    1195 .......................AATAGGGCAGATGGTATAGGGATATG 1220
                                |||         |||:::   |||||
     507 IleLysValAspTyrMetLeuLysAsnGlyLeuGlyGlyValMetIleTr 523
```

```
    1221 GCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATA 1270
         |:::::          ::::::      |||:::::      :::
     523 pGluIleThrAlaAspArgLysProGlyThrAsnAspHisProLeu.... 538

    1271 TTTTTCTTACAAATATTTGGAAACATTTAAAT..................... 1302
             |||   ::::::   :::||||||:::
     539 .....LeuAspThrValLeuGlnHisLeuGlyGluLysProProAlaTrp 553

    1303 ...CCTGAAGTA..............CAAACTCCAAAAGACCTTACTAT 1334
           |||::::::             :::   |||:::::::::|||::
     554 IleProAspThrTyrTyrIleGlySerAsnIleProSerAsnIleThrVa 570

    1335 AACTGAA............................AACCCTGAAGACT 1354
         :   |||                          :::||||||||||
     570 lProGluProThrProLeuProProSerAsnGluThrThrProGluAspA 587

    1355 GTAGCACAATA..................................... 1365
           :::|||
     587 snGlnThrAsnProAsnProSerGlnGlyAsnGluThrAsnProAsnPro 603

    1366 .........GATGAATATGTTCCA......GGACTCGTTATTCCAACCAT 1400
                 :::|||   :::|||      :::  :::   |||:::
     604 SerProGlyAsnGluThrThrProSerAspAsnGlnThrThrProSerTh 620

    1401 AGGGATATATTACAAACACAATGAT..............GCTATATGGA 1435
          |||   :::   |||   ::::::               :::   |||
     620 rGlyAspPheValLysProGlySerLeuSerValLysValThrAspTrpG 637

    1436 AAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGG 1485
           :::   :::|||::::::                     |||:::|||
     637 lyAsnThrGluTyrAspValThrLeuAsnLeuGlyGlyThrTyrAspTrp 653

    1486 GACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCA 1535
           :::||||||||||         |||:::   ||||||:::::::::
     654 ...ValValLysVal.........LysLeuLysAspGlySerSerValSe 666

    1536 TTATTATAACACTGACTATAAA......GAAAGCTCTATTATTATATGGA 1579
             :::   :::      |||      |||::::::   :::::::::
     666 rSerPheTrpSerAlaAsnLysAlaGluGluGlyGlyTyrValValPheT 683

    1580 AAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCT 1620
                                  |||:::::::||||||
     683 hrProValSer..............TrpAsnArgGlyPro 691

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB84784

seq_documentation_block:
ID   AAB84784 standard; Protein; 1215 AA.
XX
AC   AAB84784;
XX
DT   11-JUL-2001  (first entry)
XX
DE   Pyrococcus heat resistant chitinase.
XX
KW   Chitinase; heat resistant; disease.
XX
OS   Pyrococcus sp. KOD-1.
XX
PN   JP2001054381-A.
XX
PD   27-FEB-2001.
XX
PF   13-AUG-1999;   99JP-0229517.
XX
PR   13-AUG-1999;   99JP-0229517.
XX
PA   (BEAB-) BE ABLE KK.
XX
DR   WPI; 2001-310635/33.
DR   N-PSDB; AAF87608.
XX
PT   A highly heat-resistant chitinase -
XX
PS   Claim 3; Page 13-16; 19pp; Japanese.
XX
CC   The present invention relates to a highly heat-resistant chitinase
CC   maintaining at least 90% chitinase activity after treated at 100
CC   degrees C for 1 minute. The chitinase can be used for the creation
CC   of plants resistant to diseases. The present sequence is the
CC   chitinase of the invention.
XX
SQ   Sequence   1215 AA;


alignment_scores:
            Quality:  148.00            Length:    647
              Ratio:  0.576               Gaps:     28
 Percent Similarity:  39.722   Percent Identity: 17.311

alignment_block:
 US-09-579-383-2 x AAB84784   ..

 Align seg 1/1  to: AAB84784  from: 1  to: 1215

     160 CATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGG 209
         ||||||||||:::|||::::::   :::                    ::
     139 HisLeuLysAlaAsnThrThrTyrTyrTyrGlyValValProValLeuAl 155

     210 AGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAA 259
         :||||||
     155 aAspGly..................................... 157

     260 ATGATAGAAAATCACCAAGACAAATT.................TTAGAG 291
           :::|||   ||||||   ::::::                 ||||||
     158 ..SerArgGlySerProSerAsnValLeuAlaIleThrThrProLeuGlu 173

     292 GAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATG 341
           |||:::                  :::      ||||||   |||||
     174 ProTyrArg.................ValIleValTyrTyrIleSerTr 184

     342 GAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGG 391
         |:::     :::       |||     :::::::::   |||
     184 pGlyArgTyrAla......ArgLysPheTyrValSerAspIleProTrpG 199

     392 TGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTA 441
           ::::::   ::::::   :::|||   :::::::::      |||
     199 luLysValThrHisValAsnTyrAlaPheLeuAspLeuLysGluAspGly 215

     442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
         :::      |||
     216 ThrValAlaPhe..................................... 219

     492 AGAATATGAAACCTATGGTATGATGCTTAAT......GAAATTAGACGTA 535
           |||:::||||||:::       ||||||         :::::::::
     220 ....TyrAspThrTyrAlaAspProLeuAsnLeuGluAlaMetLysGluT 235

     536 TCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAA 585
           ::::::      |||   |||   :::|||:::|||:::||||||
     235 yrLysArgLysTyrProAlaValLysValLeuIleSerValGlyGlyTrp 251

     586 ACC.....................TATATGATAGATATAGAAAAAGAAAT 614
         |||                               |||      |||:::
     252 ThrLeuSerLysTyrPheSerValValAlaAlaAspProAlaLysArgGl 268

     615 TGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATG 664
           :::::::::      ::::::::::::::::   ::::::|||||||
     268 nArgPheAlaGluThrAlaIleGluIleLeuArgLysTyrAsnLeuAspG 285

     665 GTGTAGATATTGACTGGGAA.........CCACATGGGAAGTTTTACAAC 705
         ||:::|||||||||||||||               |||         |||
     285 lyIleAspIleAspTrpGluTyrProGlyGlyGlyGlyMetAlaGlyAsn 301

     706 TTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTT 755
           ::::::   :::   ::::::   ::::::   |||::::::   ||
```

```
 302 TyrGluSerProAspAspGlyLysAsnPheValLeuLeuLeuLysAspLe 318

 756 AAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAA 805
     ||||::::::::::   :::
 318 uArgGluAlaLeuAspLysAla..............................  325

 806 ATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAA 855
                                        :::|||::::::
 326 ....................................AlaLysGluAsp  329

 856 GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAA 905
     ::::::   |||
 330 HisLysAspTyr.....................................  333

 906 AGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATA 955
                           :::|||:::|||:::|||      :::
 334 .....................LeuLeuThrAlaAlaThrProAlaAspP 343

 956 TTTTTAATACA.....................GCAAAGGAGAAAATAGAT 984
           ::::::                    |||:::::::   :::|||
 343 roValLysAlaGlyArgIleAspTrpValGluAlaSerLysTyrLeuAsp 359

 985 CTTGTATTTATTCAAACATAC............................. 1005
       :::   |||:::||||||
 360 SerIleAsnIleMetThrTyrAspTyrHisGlyAlaTrpGluThrIleTh 376

1006 ...............................AATTTAGAAACTACAAATC 1024
                                    |||         |||:::
 376 rGlyHisLeuAlaProLeuTyrCysAspProAsnAlaProTyrThrAspG 393

1025 CAGATATA.......................ATGGTAGATATGTACTTA 1050
       ::::::                      |||:::   |||:::
 393 luAsnValLysTyrHisPheCysValAsnTyrThrValGlnTrpTyrIle 409

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
     :::|||:::           |||   :::|||||||:::   |||   ||
 410 GlnHisVal......ProAspLysThrLysIleThrValGlyLeuProPh 424

1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTAT 1150
     |      :::::::::   ::::::      |||||||||:::   |||
 424 e...TyrSerArgSerPheAlaAsnValProProGluAsnAsnGlyLeuT 440

1151 TAGAATTGGTAGGAAAAACA.............................. 1170
       :::      :::   |||
 440 yrGlnProPheSerGlyThrProAlaGlyThrTrpGlyProAlaTyrGlu 456

1171 ............................ATACATGATAAAAATCAAAATAA 1193
                                  :::   :::|||||||||:::::
 457 ThrTyrGlyValMetAspTyrTrpAspValAlaGluLysAsnGlnSerSe 473

1194 T.................................................. 1194
      :
 473 rGluTyrGluTyrHisTrpAspProIleAlaGlnValAlaTrpLeuTyrS 490

1194 .................................................. 1194

 490 erProSerLysArgIlePheIleThrPheAspAspProArgAlaIleGly 506

1195 .......................AATAGGGCAGATGGTATAGGGATATG 1220
                            |||         |||:::   |||||
 507 IleLysValAspTyrMetLeuLysAsnGlyLeuGlyGlyValMetIleTr 523

1221 GCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATA 1270
     |::::::         ::::::      |||::::::      :::
 523 pGluIleThrAlaAspArgLysProGlyThrAsnAspHisProLeu.... 538

1271 TTTTTCTTACAAATATTTGGAAACATTTAAAT................... 1302
         |||   ::::::   :::||||||:::
 539 .....LeuAspThrValLeuGlnHisLeuGlyGluLysProProAlaTrp 553

1303 ...CCTGAAGTA..............CAAACTCCAAAAGACCTTACTAT 1334
        |||::::::               :::   |||:::::::::|||::
 554 IleProAspThrTyrTyrIleGlySerAsnIleProSerAsnIleThrVa 570

1335 AACTGAA............................AACCCTGAAGACT 1354
       :   |||                         :::|||||||||
 570 lProGluProThrProLeuProProSerAsnGluThrThrProGluAspA 587

1355 GTAGCACAATA..................................... 1365
       :::|||
 587 snGlnThrAsnProAsnProSerGlnGlyAsnGluThrAsnProAsnPro 603

1366 .........GATGAATATGTTCCA......GGACTCGTTATTCCAACCAT 1400
             :::|||   :::|||      :::   :::   |||:::
 604 SerProGlyAsnGluThrThrProSerAspAsnGlnThrThrProSerTh 620

1401 AGGGATATATTACAAACACAATGAT..............GCTATATGGA 1435
      |||   :::   |||   ::::::               :::   |||
 620 rGlyAspPheValLysProGlySerLeuSerValLysValThrAspTrpG 637

1436 AAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGG 1485
       :::   :::|||::::::                    |||:::|||
 637 lyAsnThrGluTyrAspValThrLeuAsnLeuGlyGlyThrTyrAspTrp 653

1486 GACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCA 1535
       :::||||||||||         |||:::   ||||||:::::::::
 654 ...ValValLysVal.........LysLeuLysAspGlySerSerValSe 666

1536 TTATTATAACACTGACTATAAA......GAAAGCTCTATTATTATATGGA 1579
       :::   :::      |||      |||:::::::   :::::::::
 666 rSerPheTrpSerAlaAsnLysAlaGluGluGlyGlyTyrValValPheT 683

1580 AAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCT 1620
                                 |||::::::||||||
 683 hrProValSer..............TrpAsnArgGlyPro 691
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18255

seq_documentation_block:

ID AAB18255 standard; Protein; 1817 AA.
XX
AC AAB18255;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:112.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 270-274; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded

```
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   1817 AA;


alignment_scores:
            Quality:  144.00              Length:    558
              Ratio:  0.514                Gaps:     31
 Percent Similarity:  50.179   Percent Identity: 20.968

alignment_block:
 US-09-579-383-2 x AAB18255   ..

 Align seq 1/1  to: AAB18255  from: 1  to: 1817

      64 ACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGA 113
         |||   :::||||||:::::::::|||:::
     355 ThrLysGlnGlyLysHisAspMetAsnAsp.................Ty 365

     114 AAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATC 163
          |||   ||||||   ::::::   :::|||:::   ::::::::::::
     365 rAsnMetAsnLysLysAsnAsnMetAspIleAsnIleThrIleAsnAsnA 382

     164 TTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT 213
           ::::::||||||:::|||   :::            |||      :::
     382 snAsnAsnAsnAsnAsnLysIle............TyrAsnAspAsn 394

     214 GGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAA..... 258
              |||      :::   :::|||:::   ||||||::::::
     395 ThrLeuAsn......ValTyrAsnAsnSerTyrAsnIleHisSerAsnHi 409

     259 .......AATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAA 301
              ||||||:::::::::      |||:::||||||::::::   |
     409 sLeuMetAsnAspLysArgLysAsnAlaGlnValLeuGluLysHisLeuL 426

     302 AAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAA 351
         ||           :::::::::      :::      :::|||   :::
     426 ys...........MetLeuCysAspAsnPhe......PheAsnLeuGlu 436

     352 GGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTT 401
               :::::::::::::|||   :::|||      ::::::
     437 GluPheTyrSerSerAsnIleIleIleAsnAsnMetAspIleGluTyrTh 453

     402 ATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCAT 451
          |||      |||            :::||||||:::      :::   |
     453 rTyrAspTyrPhe...........IleLeuTyrGlu......LysCysP 464

     452 TTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAA 501
         ||         :::|||::::::                 ::::::|||
     464 heLeuProIleGluArgIleVal...........HisValAsnTyrMet 476

     502 ACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCC 551
            |||   :::      |||:::   |||:::   :::|||:::|||
     477 LysTyr...LeuTyrLysAsnAsnGluArgLysLysAsnLysIleArgLy 492

     552 AGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATA 601
             :::|||   ||||||:::               |||      ||||
     492 sPheLeuIleThrLeuLeuGlu.............TyrSerArgAspI 504

     602 TAGAAAAAGAAATTGATTATGTGGATAAAATATTG...AAGCTTGTTAAT 648
         ||:::   ::::::         :::   ||||||   |||      |||
     504 leAsnPheHisPhePheIlePheAsnLeuIleLeuTyrLysCysLysAsn 520

     649 GATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTT 698
         :::|||      :::      ::::::      :::   :::      ||
     521 GluPheProCysSerIlePheGluLeuHisIleSerGlnTyrLeuTyrPh 537

     699 TTAC...AACTTAAATGAATTAAATTTTTCAAATTATTATATT....... 738
         |:::   :::||||||||||||||||:::::::::   ||||||
     537 ePheValLysLeuAsnGluLeuAsnIleLysAspAlaTyrIleTyrTyrP 554

     739 ......................AAATTAATTAACTTGTTAAGAAAAACT 765
                                    :::|||   :::   ||||||:::
     554 heAsnAsnPheLysTyrGlnAspMetIleIleTyrPheSerArgLysAla 570

     766 ATT...CCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGC 812
         :::   |||   |||         :::::::::   :::::::::   ::
     571 PheTyrProTrpGlu......ThrAsnValGluGlnGlnLysLysGlnTh 585

     813 ATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTC 862
         :||||||   :::
     585 rLeuSerTyrIle..................................... 589

     863 CATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTA 912
           ||||||   |||                  |||:::   ||||||:::
     590 ..TyrAsnAspLys..............IleLysLysAsnLysLys... 599

     913 CATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAAT........ 954
         :::::::::      ::::::   :::|||::::::|||
     600 AsnAsnSerTyrTyrGluMetAsnAsnAsnThrTyrMetAsnGluHisGl 616

     955 ......................ATTTTTAATACAGCAAAGGAGAAAATAG 982
                                    :::|||      :::::::::::::
     616 yTyrThrAspIleGluAsnGluArgLeuAsnLysLysAsnLysArgLeuA 633

     983 ATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAATCCAGATATA 1032
         :::::                  :::      :::|||      ||||||
     633 snVal..............ArgGlyArgThrAsnThrLeuAspAspIle 644

    1033 ATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATATAACAT 1082
         :::|||:::         :::|||                ||||||||
     645 IleValSer.........AspHisGlyAsnSerTyrAspLysTyrAsnTh 658

    1083 CACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTC 1132
          :::                  :::||||||||
     658 rSer.....................LysHisAsnArg............. 663

    1133 CCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAA 1182
               :::|||:::   ::::::   :::   |||      ::::::|||
     664 .....ArgLysAsnHisIleAsnGluMetLysLysLysGlnAsnAsnLys 678

    1183 AATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTTAT 1232
         ::::::|||:::      :::||||||      :::      :::
     679 LysLysAsnThrLeuPheValAspGlyLysAspMetGluGlyIleGlyLy 695

    1233 GAAAGAACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAA 1282
          :::::::::      :::   :::::::::   :::||||||   :::|
     695 sGluLysGluLysGluAsnLysAsnMetAsnAsnAsnIlePheTyrAsnA 712

    1283 AT......................ATTTGGAAACATTTAAATCCTGAA 1308
         ||                              ::::::::::::|||   :::
     712 snSerTyrSerAsnIleAsnAsnSerSerTyrSerAsnIleAsnAsnAsp 728

    1309 GTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAG 1358
         :::   :::      ::::::|||   ::::::|||   :::      ||
     729 IleTyrSerValAspAsnMetThrSerValAsnAsnThrLysTyrValSe 745
```

```
 1359 CACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATAT 1408
 745 rGlyValProSerTyrAlaHisValLeuIleAsnLysGlnValAsnGluT 762

 1409 ATTACAAA...........CACAATGATGCTATATGGAAAACTAGATCT 1446
 762 yrTyrGlnGlyLeuProAsnTyrAsnAsnMetMetIleLys...GlySer 777

 1447 TATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAA 1496
 778 HisIleIleAsnGluLeuProLysAsnAsnTyr..... 788

 1497 AGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCATTAT..... 1539
 789 ....IleTyrGluAsnAsnTyrIleGlyGln......AsnTyrLeuMetT 802

 1540 ..........TATAACACTGACTATAAAGAAAGCTCTATTATTATATGG 1578
 802 hrAsnProLeuTyrAsnLysGluThrLysAspIlePheTyrThrIleTyr 818

 1579 AAAGGGGAACCATATTTAATTAAA 1602
 819 Lys.........TyrLeuPheLys 823
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR64823

seq_documentation_block:

ID AAR64823 standard; Protein; 561 AA.
XX
AC AAR64823;
XX
DT 23-JUL-1995 (first entry)
XX
DE Chitinase A.
XX
KW Chitinase A; protein secretion; biological control agent;
KW transgenic plant; chitin; plant pathogen; nematode; fungus;
KW crop protection; pCHIT1251.
XX
OS Serratia marcescens QMB1466.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= Sig_peptide
XX
PN US5374540-A.
XX
PD 20-DEC-1994.
XX
PF 26-MAR-1984; 84US-0593691.
XX
PR 26-MAR-1984; 84US-0593691.
PR 18-JUL-1986; 86US-0888033.
PR 09-JUL-1990; 90US-0550253.
XX
PA (DNAP) DNA PLANT TECHNOLOGY CORP.
XX
PI Jones JDG, Suslow T;
XX
DR WPI; 1995-035648/05.
DR N-PSDB; AAQ76290.
XX
PT Plant cell contg. signal sequence of chitinase A gene - for
PT secretion of foreign polypeptide, esp. chitinase A for protection
PT against pathogenic fungi and nematodes.
XX
PS Disclosure; Fig. 1A-1C; 20pp; English.
XX
CC Serratia marcescens QMB1466 genomic DNA was partially digested,
CC ligated into vector pLAFR1, packaged into lambda phage particles,
CC and used to transfect Escherichia coli DH1 (ATCC 33849) cells.
CC Plating on chitin medium was used to screen for chitinase activity.
CC Clone C12 (ATCC 67152) was isolated. Its sequence is given in
CC AAQ76290, and the deduced protein sequence in AAR64823.
XX
SQ Sequence 561 AA;

alignment_scores:
Quality: 142.50 Length: 425
Ratio: 0.720 Gaps: 20
Percent Similarity: 46.588 Percent Identity: 20.471

alignment_block:
US-09-579-383-2 x AAR64823 ..

Align seg 1/1 to: AAR64823 from: 1 to: 561

```
  25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
  11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

  75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
  27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

 116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
  44 spGlnAlaAlaThr.....AlaTyrAsnAsnLeuValLysVal 56

 166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
  57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

 214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA..... 246
  73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

 247 .....AATATAAATAAAAATGAT 264
  90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

 265 AGA..... 267
 107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

 267 ..... 267

 123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

 268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
 140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

 313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
 155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

 354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
 171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

 404 ATATTGCATTTGCTCGCATT.....AATATGTTATATGAT 438
 186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

 439 .....GTATCTAGACCATTTAA 455
 203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

 456 TGGAAGACAAAGATTCCTATTAAGAAAACAC.....GGCTTAG 493
 219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

 494 AATATGAAACCTATGGTATGATG..... 516
```

```
      236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

      517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
          ::::::::::::      ::::::::::::::::|||||||:::   |||||
      253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

      567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
          |   |||:::||||||   |||                   :::|||   |
      269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

      599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
          ||      |||:::      :::|||   ::::::   ::::::::::::
      286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

      649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
             :::      :::||||||||||||||||||||||||   |||   ||
      303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

      693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
          ||||            ||||||::::::   :::   ::::::   |||:
      319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

      737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
          ::   |||::::::   ||||||      :::   :::
      336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

      787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                  |||||||:::   :::|||
      349 ..........................LeuSerAlaGluThrGly..... 354

      837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                             |||:::   |||:
      355 ................................ArgLysTyrGluLeuT 360

      884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
          :::::   |||::::::::::|||:::          :::|||
      360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

      934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                 :::|||:::|||:::::::::::::||
      375 ........................TyrAsnValAlaGlnAsnSerMetAs 383

      984 TCTTGTATTTATTCAAACATACAAT 1008
          |   :::|||::::::::::|||:::
      383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:AAW09402

seq_documentation_block:
ID   AAW09402 standard; Protein; 561 AA.
XX
AC   AAW09402;
XX
DT   19-FEB-1997  (first entry)
XX
DE   Serratia marcescens Chitinase A.
XX
KW   Chitinase; ChiA; chitin degradation; transgenic plant; tomato;
KW   cold resistance; tolerance; frost damage; reducing sugar content;
KW   increase; ATCC 9901; ds.
XX
OS   Serratia marcescens QMB 1466.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= signal_peptide
FT                   /note= "predicted cleavage site is between residues
FT                          23 and 24"
FT   Protein         24..561
FT                   /label= Chitinase_A
FT                   /note= "predicted mature sequence"
XX
PN   US5554521-A.
XX
PD   10-SEP-1996.
XX
PF   26-MAR-1984;   84US-0593691.
XX
PR   19-DEC-1994;   94US-0358901.
PR   26-MAR-1984;   84US-0593691.
PR   18-JUL-1986;   86US-0888033.
PR   09-JUL-1990;   90US-0550253.
PR   14-AUG-1992;   92US-0930970.
XX
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
XX
PI   Jones JDG,  Suslow TV;
XX
DR   WPI; 1996-424655/42.
DR   N-PSDB; AAT49479.
XX
PT   Recombinant plant with increased resistance to cold - and increased
PT   reducing sugar content, contains DNA encoding chitinase, partic. to
PT   protect tomatoes against frost damage.
XX
PS   Example 3; Fig 1A-C; 25pp; English.
XX
CC   The present sequence is the Chitinase A enzyme encoded by the 2.3 kb
CC   insert of plasmid pCHIT1251 which was isolated from genomic DNA of
CC   Serratia marcescens QMB 1466 (ATCC 9901). The chiA gene can
CC   be used for transforming plants, esp. tomatoes, to give increased
CC   resistance to frost damage and to produce sweeter fruit.
XX
SQ   Sequence   561 AA;


alignment_scores:
           Quality:  142.50              Length:    425
             Ratio:  0.720                Gaps:     20
 Percent Similarity:  46.588   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x AAW09402   ..

 Align seg 1/1  to: AAW09402  from: 1  to: 561

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||::::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||                :::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   :::   :::|||      |||:::|||
      73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ..............................AATATAAATAAAAATGAT 264
                                        ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA........................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 ........................................... 267
```

```
123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
        |||:::|||      :::|||||||:::   |||      ||||||
140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
    :::         ::::::::::::|||:::   :::|||:::      ||
155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
    |        :::      :::|||:::   |||      ::::::   :
171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
    :::::   ::::::   |||               |||   :::   |||
186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

439 ..............................GTATCTAGACCATTTAA 455
                                 ::::::|||      ::
203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
    :|||||||::   |||   :::      |||            :::|||:
219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

494 AATATGAAACCTATGGTATGATG.......................... 516
    ::          |||:::
236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
    ::::::::::::      ::::::::::::::::||||||:::   |||||
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
    |   |||:::||||||   |||                 :::|||   |
269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
    ||      |||:::      :::|||  ::::::   ::::::::::::
286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
       :::      :::||||||||||||||||||||||||   |||   ||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
    ||||            ||||||::::::   :::   ::::::   |||:
319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
    ::   |||::::::   ||||||      :::   :::
336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                              ||||||:::   :::|||
349 ..........................LeuSerAlaGluThrGly..... 354

837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                        |||:::   |||:
355 ................................ArgLysTyrGluLeuT 360

884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
    :::::   |||:::::::::|||:::          :::|||
360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                          :::|||:::|||:::::::::::::||
375 ......................TyrAsnValAlaGlnAsnSerMetAs 383

984 TCTTGTATTTATTCAAACATACAAT 1008
    |   :::|||:::::::::|||:::
383 pHisIlePheLeuMetSerTyrAsp 391
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW29178

seq_documentation_block:
ID   AAW29178 standard; Protein; 561 AA.
XX
AC   AAW29178;
XX
DT   06-MAR-1998  (first entry)
XX
DE   S. marcescens chitinase A.
XX
KW   Serratia marcescens; chitinase A; transgenic plant; cold resistant;
KW   reducing sugar; increased level.
XX
OS   Serratia marcescens.
XX
PN   US5633450-A.
XX
PD   27-MAY-1997.
XX
PF   01-DEC-1995;   95US-0566347.
XX
PR   19-DEC-1994;   94US-0358901.
PR   26-MAR-1984;   84US-0593691.
PR   18-JUL-1986;   86US-0888033.
PR   09-JUL-1990;   90US-0550253.
PR   14-AUG-1992;   92US-0930970.
PR   01-DEC-1995;   95US-0566347.
XX
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
XX
PI   Jones JDG,  Suslow TV;
XX
DR   WPI; 1997-297423/27.
DR   N-PSDB; AAT86836.
XX
PT   Transgenic plants expressing chitinase - with increased cold
PT   resistance or sugar content
XX
PS   Example 3; Columns 25-30; 25pp; English.
XX
CC   This sequence represents the Serratia marcescens chitinase A.
CC   Information from the DNA coding it (AAT86836) was used in the creation of
CC   constructs which led to the expression of the bacterial chitinase gene in
CC   plant cells. Plants which have been transformed with chitinase-encoding
CC   DNA so that they express active chitinase are resistant to cold damage or
CC   have enhanced reducing sugar levels. The plants, especially tomato
CC   plants, have improved frost resistance or produce sweeter fruits.
XX
SQ   Sequence   561 AA;


alignment_scores:
            Quality:  142.50            Length:    425
              Ratio:  0.720               Gaps:     20
 Percent Similarity:  46.588   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x AAW29178   ..

 Align seg 1/1  to: AAW29178  from: 1  to: 561

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||               ::::::::::::      :::
```

```
 44 spGlnAlaAlaThr.....AlaTyrAsnAsnLeuValLysVal 56

166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
    |||:::      :::      |||:::::::::   :::   ||||||
 57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
     |||   :::   :::|||      |||:::|||
 73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

247 ............................AATATAAATAAAAATGAT 264
                                :::::||||||:::
 90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

265 AGA.......................................... 267
    |||
107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

267 ........................................... 267

123 rAspAlaThrGluIleValValAlaAspThrAspGlyArgHisLeuProP 140

268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
         |||:::|||      :::||||||:::   |||      ||||||
140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
    :::         ::::::::::::|||:::   :::|||:::      ||
155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
    |         :::      :::|||:::   |||      ::::::   :
171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
    :::::   ::::::   |||               |||   :::   |||
186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

439 ............................GTATCTAGACCATTTAA 455
                               ::::::|||      ::
203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
    :||||||:::   |||   :::      |||           :::|||:
219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

494 AATATGAAACCTATGGTATGATG....................... 516
    ::          |||:::
236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
    ::::::::::::      ::::::::::::::|||||||:::   |||||
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
    |   |||:::||||||   |||                 :::|||   |
269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
    ||      |||:::      :::|||  ::::::   ::::::::::::
286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
       :::      :::||||||||||||||||||||||   |||   ||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
    ||||           ||||||::::::   :::   ::::::   |||:
319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
    ::   |||::::::   ||||||      :::   :::
336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                  ||||||:::   :::|||
349 ..........................LeuSerAlaGluThrGly..... 354

837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                              |||:::   |||:
355 ................................ArgLysTyrGluLeuT 360

884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
    :::::   |||::::::::::|||:::           :::|||
360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374

934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                          :::|||:::|||:::::::::::||
375 ......................TyrAsnValAlaGlnAsnSerMetAs 383

984 TCTTGTATTTATTCAAACATACAAT 1008
    |   :::|||:::::::::|||:::
383 pHisIlePheLeuMetSerTyrAsp 391
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW60179

seq_documentation_block:

```
ID   AAW60179 standard; Protein; 561 AA.
XX
AC   AAW60179;
XX
DT   03-SEP-1998  (first entry)
XX
DE   Serratia marcescens chitinase A sequence.
XX
KW   Chitinase A; pathogen; plant protection; Serratia marcescens; E. coli;
KW   Pseudomonas; Erwinia; frost damage resistance; plant pathogen; fungus;
KW   nematode.
XX
OS   Serratia marcescens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /note= "putative signal peptide"
FT   Cleavage-site   23..24
FT                   /note= "predicted cleavage site"
FT   Protein         24..561
FT                   /note= "mature protein"
XX
PN   US5776448-A.
XX
PD   07-JUL-1998.
XX
PF   01-AUG-1996;   96US-0693835.
XX
PR   19-DEC-1994;   94US-0358901.
PR   26-MAR-1984;   84US-0593691.
PR   18-JUL-1986;   86US-0888033.
PR   09-JUL-1990;   90US-0550253.
PR   14-AUG-1992;   92US-0930970.
PR   01-AUG-1996;   96US-0693835.
XX
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
XX
PI   Jones JDG,  Suslow TV;
XX
DR   WPI; 1998-397941/34.
DR   N-PSDB; AAV35601.
XX
PT   Use of heterologous chitinase nucleic acid - for transforming
PT   bacteria or plants, for increasing plant resistance to plant
PT   pathogens, increase chilling resistance and increasing sweetness.
XX
PS   Example 3; Fig 1A-C; 25pp; English.
XX
CC   This represents a Serratia marcescens chitinase A. The invention
```

```
CC   provides a method for the production of a bacterial cell capable of
CC   producing chitinase which comprises transforming the bacterial cells
CC   with a nucleic acid encoding the chitinase, the nucleic acid being
CC   isolated from a heterologous source, and whereby the bacterial cell
CC   is capable of expressing the nucleic acid. The bacterial cell is selected
CC   from Pseudomonas sp., E. coli or Erwinia sp. The products can be used to
CC   enhance plant growth by biological control of plant pathogens such as
CC   fungi, nematodes, insects and disease agents.  Plants transformed with
CC   the chitinase DNA can also have resistance to frost (freezing) damage
CC   or chilling damage, increased levels of reducing sugars or sweetness
CC   in fruits or plants and enhanced post-harvest storage life.  The
CC   products can also be used for the production of chitinase for use as an
CC   antibiotic.
XX
SQ   Sequence   561 AA;


alignment_scores:
           Quality:  142.50              Length:    425
             Ratio:  0.720                Gaps:     20
 Percent Similarity:  46.588   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x AAW60179   ..

 Align seg 1/1  to: AAW60179  from: 1  to: 561

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||              :::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   :::   :::|||      |||:::|||
      73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ............................AATATAAATAAAAATGAT 264
                                     ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA........................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 ........................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
              |||:::|||      :::|||||||:::   |||.     ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::         :::::::::::|||:::   :::|||:::      ||
     155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         :::      :::|||:::   |||      ::::::   :
     171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
         :::::   ::::::   |||               |||   :::   |||

     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

     439 ..............................GTATCTAGACCATTTAA 455
                                      ::::::|||      ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
         :|||||||:::   |||   :::      |||            :::|||:
     219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

     494 AATATGAAACCTATGGTATGATG.......................... 516
         ::           |||:::
     236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::::|||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
         |   |||:::||||||   |||                  :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
            :::      :::||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluProProGlyGl 319

     693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
         ||||             ||||||::::::   :::   ::::::   |||:
     319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

     737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
         ::   |||::::::   ||||||      :::   :::
     336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                      ||||||:::   :::|||
     349 ..........................LeuSerAlaGluThrGly..... 354

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                           |||:::   |||:
     355 ..............................ArgLysTyrGluLeuT 360

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         :::::   |||:::::::::|||:::             :::|||
     360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                  :::|||:::|||:::::::::::||
     375 ......................TyrAsnValAlaGlnAsnSerMetAs 383

     984 TCTTGTATTTATTCAAACATACAAT 1008
         |   :::|||:::::::::|||:::
     383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18180

seq_documentation_block:
ID   AAB18180 standard; Protein; 2295 AA.
XX
AC   AAB18180;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:37.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
```

XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 92-98; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 2295 AA;

```
alignment_scores:
            Quality:  139.50              Length:    731
              Ratio:  0.431                Gaps:     37
 Percent Similarity:  44.323   Percent Identity: 18.741

alignment_block:
 US-09-579-383-2 x AAB18180   ..

 Align seg 1/1  to: AAB18180  from: 1  to: 2295

       1 ATGAATTTTAAAATATCAATATTTTTAATTATAGTATCCATC..... 42
         :::||||||   :::|||:::||||||:::|||:::::
       8 LeuAsnPheIlePheSerPhePheLeuLeuIleIleLysCysAspGluSe 24

      43 .....TTGTATTCTGCAAATTCCAGAACCTTGA 70
                        :::|||      :::::::::   ||||
      24 rValSerAsnGlyArgLysGluIleTyrPheAspAspAspGluLysLeuL 41

      71 AA.................GGAAAAAATAATATAAATAATTCATTGGGA 102
         ||                  ::::::|||||||||   ::::::|||
      41 ysLeuSerSerPhePheAspArgSerThrAsnIleAsnLeuAspValGly 57

     103 .....ATAATACGGGAA......AATAAAAA 122
                           :::   ||||||      :::|||::

      58 GluAsnAspGluLeuSerSerTyrValProArgGluValAspGluLysLy 74

     123 TAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTTAAATCGA 172
         :|||   ::::::   :::|||   :::|||   ::::::   ||||||:
      74 sLysLysAsnLysLysAspIle...AspSerLysGluAsnSerLysSerG 90

     173 ATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGTAAC 222
         ::|||:::   :::                                |||
      90 lyAsnAsnIleTyr.............................Asn 95

     223 TCTAGAATTACAAAAAATAATAAAAATATAAATAAAAAT.......... 261
         :::      ||||||||||||:::::::::|||   |||
      96 LysAspAsnThrLysAsnAsnGluAspValAsnTyrAsnValValLeuLy 112

     262 .GATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAAC 310
          |||   ::::::   :::   |||:::   :::   :::||||||::::
     112 sAspGlyArgAlaLysGluGlyIleIleThrAsp...GluLysArgArgS 128

     311 AAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGA 360
         :::::         |||         :::   |||::::::      :::
     128 erSerThrLysAspGlyLysAsnLysGluGlnAsnAsnAsnLysMetAsn 144

     361 GCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGC 410
         :::      :::   |||:::|||
     145 SerAspAspValHisAspAsnAsn..................... 152

     411 ATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGAA 460
                    |||   :::   |||:::::
     153 .............AsnAsnMetAsnAspIleAsn............... 159

     461 GACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGT 510
               |||:::::::::                          |||:::
     160 ........PheValValGlu.....................TyrAsn 165

     511 ATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAAT 560
            |||:::::::::      :::|||               |||   ::
     166 LysMetIleAspAsnTyrAspLysIleLeu............AspGluLe 178

     561 TATTCTTTTATCCTTAGGTGGAGAAACCTAT...........ATGATAG 598
         :||||||   |||::::::   ::::::|||           |||:::|
     178 uIleLeuLysSerIleAsnArgAsnAsnTyrAsnTyrPheAsnMetLeuA 195

     599 AT.....................ATAGAAAAAGAAATT...GATTATGTG 624
         ||                  ::::::||||||:::   |||   :::
     195 spGluTyrSerLeuGlnThrLysLeuAsnLysGluMetTyrAspSerLeu 211

     625 GATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATAT 674
         :::   :::::::::|||:::|||:::   :::                ||
     212 AsnTyrLeuIleArgLeuMetAsnAsnLysAsnSerArgLysTyrPheIl 228

     675 TGACTGGGAACCACATGGGAAGTTTTACAACTTA...AATGAATTAAATT 721
         |:::::::::   :::   |||         :::   |||:::::::|||
     228 eSerPheSerAsnAsnGluLysLysLysIleIleLysAsnAspMetAsn. 244

     722 TTTCAAATTATTATATTAAA......TTAATTAACTTGTTAAGAAAAACT 765
           :::|||   ||||||:::      :::::::::|||:::|||
     245 ..GluAsnIleTyrIleArgHisPheIleValSerLeuPheArg...Trp 259

     766 ATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATT 815
               :::   |||||||||:::         ::::::|||      ::
     260 TyrAsnAsnPheLysLeuIleGluThrCysPheAspLysAsnAsnPheIl 276

     816 ATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTCCAT 865
         :      :::                          ||||||::::::   |
     276 eTyrTyrIle.....................AspGluAsnLysIleT 285

     866 ATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTACAT 915
         ||:::   |||:::                    |||   :::|||
     285 yrSerTyrLysTyr...................AsnTyrLysLeu... 293

     916 AGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATAC 965
                    ::::::|||:::   :::|||:::      :::|||
     294 ...MetLeuAsnLeuPheSerSerGluAsnPheLeuTyrTyrIleAsnLe 309
```

```
 966 AGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATACAAT..... 1008
     :::|||   ::::::   |||   :::|||:::::::||||||
 309 uSerLysPheSerLeu...LeuGluIleIleAspAsnTyrAsnLysTyrS 325

1009 ........TTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050
             :::::::::   :::   |||          :::||||||:::
 325 erPheIleIleAsnAsnIleLysArgAspTyrProAsnAsnMetTyrVal 341

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
        :::   ::::::   :::   |||:::                  ::
 342 CysGlnSerPheTyrAspPheIleTyrSer.................Ty 352

1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTA. 1149
     :   |||::::::|||:::      |||      ::::::|||   |||
 352 rPheLeuSerTyrAsnHis...HisPhePheAspLysHisLysTyrLeuI 368

1150 .....TTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAAT 1194
          :::::::::   :::::::::|||:::   |||:::|||
 368 leAsnMetAspIleTrpAsnAsnSerIleGlnThrLysGlyGln...... 382

1195 AATAGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAA........ 1236
                    ||||||         |||:::   |||
 383 ..............IleGlyAsnHisLysLeuTyrLysLysLeuLysLy 394

1237 .................................................G 1237
                                                     |
 394 sLeuAsnGluAsnLeuIleLeuTyrAsnTyrIleLysAsnAspAspSerG 411

1238 AACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAAATATT 1287
     ||:::::|||      ::::::::::::   :::::::::::::|||:::
 411 luMetIleProTyrValThrLeuGluMetArgMetIlePheSerAsnPhe 427

1288 TGGAAA.......................................... 1293
        :::
 428 ThrAsnLeuLeuIleAspIleLeuAsnLysLeuTyrAsnIleAspTyrGl 444

1294 .........................CATTTAAATCCTGAAGTACAAACTC 1318
                             ::::::|||||:::   ::::::|
 444 nAspAsnIleLysGlnGluAsnValAsnValAsnProGlnArgAspAlaP 461

1319 CAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAGCACAATAGAT 1368
     ||:::|||   :::   ::::::|||      |||   |||      :::
 461 roGlnAspTyrValHisAsnLysAsnAspValAspValSerLeuLysAsn 477

1369 GAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAACA 1418
               |||       |||                    :::||
 478 ValLysGluProLysLysVal.......................GluHi 486

1419 CAATGATGCTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTG 1468
     ||||   |||:::            :::|||:::            |||
 486 sAsnLysAlaMet.........SerAsnTyrGluThrAspGluArgGlyA 500

1469 TAGACAGATATGAATGGGACTTGGTCAAAGTATGCTATGAAAAAATA... 1515
          |||:::   :::      |||      :::||||||
 500 spMetIleTyrAsp...AsnThrAsnLysGluLysPheGluLysSerGlu 515

1515 ................................................. 1515

 516 GlyThrPheAsnAsnIleSerGlyGlyGluAspThrPheLysAsnIleSe 532

1515 ................................................. 1515

 532 rGlyGlyGluAspThrPheLysAsnIleSerGluGlyAspGlyGluValA 549

1516 ........................TGCGATGGGAAAGCAGCC...... 1533
                             ||||||:::::|||
 549 spGlyAspGlyGluGlyAspGlyAspGlyAspGlyGluGlyAlaAspAsp 565

1534 ..............CATTATTATAACACTGACTATAAAGAAAGCTCTAT 1568
                   |||      :::::::|||   |||||||||:::
 566 SerSerValAspThrHisAsnAsnLysAsnAspGlyLysGluSerGluSe 582

1569 TATTATATGG..................................... 1578
          :::|||
 582 rAspValTrpAsnLeuLeuMetAspSerTyrLysLysLeuAlaAsnAspG 599

1579 ..........AAAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGA 1617
                |||   :::   |||:::::::|||
 599 luAsnPheLysLysTyrAsnLysTyrIleLeuLys............... 610

1618 CCTCCGGAAGGTCAGGCACTAGAGTCATACACAAAACTAGATGCATCCAA 1667
                      |||:::::::::   :::::::::::::::||
 611 ..............AsnLeuAspLysPheLeuAsnMetSerSerGluLy 622

1668 ATGTCCAGGGATAGAAGAATGGAATAAAAAATATCCACATAAA 1710
     |         |||:::::::::::::::::||||||      |||
 622 sLysGluAspIleAsnSerTyrLysAsnLysTyrGluLeuLys 636
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18253

seq_documentation_block:

ID AAB18253 standard; Protein; 3973 AA.
XX
AC AAB18253;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:110.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 253-263; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new

```
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   3973 AA;


alignment_scores:
            Quality:  138.00              Length:    675
              Ratio:  0.471                 Gaps:     34
 Percent Similarity:  43.407   Percent Identity: 20.889

alignment_block:
 US-09-579-383-2 x AAB18253   ..

 Align seg 1/1  to: AAB18253  from: 1  to: 3973

      1 ATGAATTTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTAT.. 48
        :::|||:::   :::::::::|||||||:::|||   :::      |||
   2770 IleAsnIleLeuValAsnLeuPheLeuPheIleAsnLysArgAspTyrAs 2786

     49 .......TCTGCAAATTCCAGAACCTTGAAAGGAAAAAATAATATAAATA 91
                ::::::   |||:::   :::::::::   ||||||   :::|
   2786 pAspGluThrCysMetSerAsnIleIleAsnAsnAspAsnAsnLysLysA 2803

     92 ATTCATTGGGAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAA 141
        ||:::   ::::::|||   |||||||||||||:::   ::::::|||:::
   2803 snLysAsnAsnLeuIle...GluAsnLysAsnGluIleTyrAsnThrAsn 2818

    142 ATACATGAGTCTTTTTCACATCTTAAA.........TCGAATAATAGTAA 182
        |||   :::|||:::::::::   |||          :::|||:::|||||
   2819 Ile...LysSerLeuLysAsnAspLysGluTyrIleAspAsnHisSerAs 2834

    183 TTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGT.............. 219
        |:::::   :::      ||||||          |||
   2834 nTyrAlaMetPhe......TyrCysAspLeuPheCysAspAspPhePheI 2849

    220 .....AACTCTAGAATTACAAAAAATAAT..................... 243
             |||::::::   :::|||:::|||
   2849 leSerAsnGlyLysLysAsnLysGluAsnValValPheHisThrLeuHis 2865

    244 ............AAAAATATAAATAAAAATGATAGAAAATCACCAAGACA 281
                   |||:::::::::|||   |||      :::   :::::
   2866 AsnMetSerHisLysGluMetSerLysTyrAspLeuIleGlyLysAsnLy 2882

    282 AATTTTAGAGGAGTAT.....................AAAAAAAGGAAAC 310
        :   ||||||:::|||                     ||||||:::|||
   2882 sTyrLeuGluAsnTyrIleAsnAsnLeuIleLeuGluLysLysLysLys. 2898

    311 AAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGA 360
                                          |||:::   :::   :::
   2899 ............................IleAsnAsnLeuAsnValHis 2905

    361 GCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGC 410
           :::      :::|||:::|||           |||||||||
   2906 IleAsnLysLysMetAspAsnAsn............IleLeuTyrSerPh 2918

    411 ATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGAA 460
         :::   ||||||||         |||   ::::::      |||
   2918 eIleAsnArgIleAsnGluThrArgAspAsnThrLysLysLysAsnLysL 2935

    461 GA......CAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACC 504
                ::::|||:::   |||:::|||   ::::::::::|||::::::
   2935 euTyrIleArgArgTyrTyrLeuLysLys...SerIleLysTyrAsnAsn 2950

    505 ......TATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACG 548
              |||:::|||
   2951 HisLeuTyrAsnMet................................... 2955

    549 TCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAG 598
              :::::::::   |||   |||         |||      :::
   2956 ....ProIlePheLeuSerLeuPheLeuArgCysValThrIleHisLeuH 2971

    599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
              |||      :::           :::||||||      |||
   2971 isTyrPheLysPheTyrAsnSerTyrIleTyrPheLeuLysHisTyrAsn 2987

    649 GATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACAT........ 690
           :::   :::                          ||||||
   2988 MetLeuHisIle.....................ProHisAlaValLe 2996

    691 ..............................GGGAAGTTTTACAACT 706
                                      :::   |||   |||:
   2996 uLysHisLeuTyrSerThrHisHisPheAsnIleAsnLeuPheValAsnM 3013

    707 TAAATGAATTA.................AATTTTTCAAATTATTATATT 738
        ::   ||||||                :::::|||||:::::
   3013 etLeuGluLeuPheTyrValPheIleGlnIleTyrAsnAsnTyrPheVal 3029

    739 AAA........................................... 741
        :::
   3030 SerPheCysAspIleSerSerCysArgAsnLysHisValGlnArgAspGl 3046

    742 .......TTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAA 784
              |||   |||      :::|||:::      :::|||||||:::
   3046 nArgCysLeuAsnAsnAsnLysAsnLysSerGluAspAsnGluLysIleT 3063

    785 TT............................................ 786

   3063 yrCysThrAsnAsnAsnGlyAspGlyTyrAspAspAspGlyTyrGlyGlu 3079

    787 ...TCAATTTCTGGT.........TCATCAAATGCTGCATTATCATGCGT 824
           ::::::||||||         ::::::|||      ::::::
   3080 LysAsnValSerGlyIleTyrLysGluAsnAsnAsnLysIleAsnValLy 3096

    825 TTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTA 874
        :::   :::   ::::::   :::   :::      |||   |||   |
   3096 sGlyAspIleTyrAsnIleAspAsnIleAsnValTyrProLeuAsnGlyL 3113

    875 AATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCA 924
        ||:::::::|||      ::::::|||   ||||||:::   :::   :::
   3113 ysLeuValSerIleTyrLeuAsnThrLeuLysGluIleLeuLysGluCys 3129

    925 GCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAG.. 972
                      |||   :::   |||      |||      |||
   3130 TyrGluCysHisLeuGlyHisMetLysAsnAsnGluAsnMetGluLysSe 3146

    973 ..........GAGAAAATAGATCTTGTATTTATTCAAACATACAATTTAG 1012
                  |||   :::   |||   |||:::      ||||||
   3146 rPhePheIleGluHisLeuLeuLeuTyrPheLeu......TyrAsnArgI 3161

   1013 AAACTACAAATCCAGATATAATGGTAGATATGTACTTATCCCATTTATAT 1062
          :::|||      :::::::::   ::::::|||::::::::::|||
   3161 leAsnThrIleTyrGluLeuPheTyrAsnPheTyrPheThrTyrLeuArg 3177

   1063 TTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAACA 1112
                ::   |||      |||:::|||   :::      |||||
   3178 ...LysLysGluAsnAsnAsnAspIleLeuLeuAspIleValAsnGluHi 3193

   1113 TAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAG 1162
        |                                 :::   :::|||:::|
   3193 s.................................IleTyrAsnLeuIleG 3199

   1163 GAAAAACAATACATGATAAA...AATCAAAATAATAATAGGGCAGATGGT 1209
        ||:::   |||:::|||:::   |||:::   ||||||      |||
   3199 lyAsnLysIleTyrAspGlnIleAsnLysIleAsnAsnPheLeuAspAsp 3215

   1210 ATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAACTGGATCATT 1259
                ::::::::::::::::::            |||   :::||
   3216 LysGlnTyrTyrTyrPheTyrIleAsn............ThrLeuThrPh 3228

   1260 CGATGTAGATATTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAAG 1309
        |
   3228 e..................................................... 3228

   1310 TACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAGC 1359
```

```
                              ||||||   |||   :::   |||
3229 .....................IleThrLeuAsnLysGlnIleCys... 3236

1360 ACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATATA 1409
                                                  |||||
3237 ..............................................IleTy 3238

1410 TTACAAACACAATGATGCTATATGGAAAACTAGATCTTATTCAATTCATG 1459
     .  |         :::    :::   :::||||||      |||   |||:::
3238 rIleIleLysLys.LysIleLeuAsnLysLeuIleTyrIleProPheIle 3254

1460 CACCTGGTGTAGACAGATATGAATGGGACTTGG...TCAAAGTATGCTAT 1506
     :::            |||   |||   ::::::   |||   |||   ::
3255 TyrHisSerLeuPheAspLysAsnLysAsnPheThrSerIleTyrHisIl 3271

1507 G......AAAAAATATGCGATGGGAAAGCAGCCCATTATT.........A 1541
     :      ::::::|||      :::|||:::   ||||||        :
3271 eAsnAsnAsnGlnTyrIleArgAsnLysAspHisIleIlePheCysSerL 3288

1542 TAACACTGACTATAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCAT 1591
     ::   :::   |||   |||   |||   ::::::   |||
3288 euIleValPheIleIleLysLeuLeuTyrValPheValLys......... 3301

1592 ATTTAATTAAATGGTGGCAACAAGGACCTCCGGAAGGTCAGGCACTAGAG 1641
                        |||         ::::::|||:::      ::
3302 .................AsnPheLysHisLysAsnValAsnPheAsnAs 3312

1642 TCATACACAAAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATGGAA 1691
     :|||   :::|||      |||   ||||||
3312 nHisGlnAsnAsnLysAsnHisGluAsnVal..................... 3322

1692 TAAAAAATATCCACATAAACCAC 1714
          |||   ::::::::::|||
3323 .....AsnThrAsnValGlyHis 3328
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB18272

seq_documentation_block:

ID AAB18272 standard; Protein; 2500 AA.
XX
AC AAB18272;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 302-309; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 2500 AA;

alignment_scores:

```
            Quality:  136.50              Length:    485
              Ratio:  0.588                 Gaps:     27
 Percent Similarity:  47.835   Percent Identity: 21.649
```

alignment_block:

US-09-579-383-2 x AAB18272 ..

Align seg 1/1 to: AAB18272 from: 1 to: 2500

```
  58 TCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAAT 107
     |||:::::::|||      ||||||:::|||:::::::::       ||
1318 SerAsnAsnLeuLeuValLysAsnGluIleThrAspLys......GluIl 1332

 108 ACGGGAAAATAAAAATAAA......ACTCATCAAACGGAAATACATGAGT 151
     |   ||||||||||||:::         ::::::   :::|||   |||:
1332 eSerGluAsnLysAsnGluLysIleIleGluAsnGluLysIleIleGluA 1349

 152 CTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCT 201
     ::   :::   :::|||      |||:::|||:::|||:::
1349 snGluLysValValLys......AsnGluAsnMetValLys......... 1360

 202 TATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATAT 251
                         |||:::::::::::::|||   |||||||||::
1361 .................AsnGluLysValValLys...AsnLysAsnVa 1370

 252 AAATAAAAATGAT.........AGAAAATCACCAAGACAAATTTTAGAGG 292
     :   ||||||:::         :::|||:::   |||   |||   :::|
1370 lValLysAsnGluAsnValValGluLysAsnSerArgPheIleLysLysG 1387

 293 AGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGG 342
     ||::::::            ::::::         |||   ::::::
1387 luHisAsnIleSerMetLeuAsnValProAsnTyrTyrGluAsnAsnThr 1403

 343 AACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGT 392
     :::::::::           ::::::   |||:::::::::|||:::||
1404 ArgGlyLysAspIleThrAsnAsnAsnAsnIleSerGlyAspProLeuVa 1420

 393 G............TCAATTTTATATATTGCATTTGCTCGCATTAATATGT 430
     |            ::::::   |||      ::::::      |||   :
1420 lAsnGlyIleSerThrLeuSerTyrLysProTyrSerThrTyrAsnCysI 1437

 431 TATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGA 480
     ::   ::::::   :::      :::      ::::::|||:::   :::
1437 leSerAsnIleIleGluGluGluLysGluIleLysLysPheValAsnLys 1453

 481 AAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATTAG 530
```

```
     |||   :::                          ||||||:::|||::
1454 LysLysAsn.....................SerLeuAsnHisIleAs 1462

 531 ACGTATCAGAAAAGTA.............CGTCCAGATGTAATTATT. 564
     :|||   :::|||:::          :::   :::   ||||||
1462 nArgAsnGluLysIleTyrIleGlyAspAsnLysLysAsnTyrIleIleG 1479

 565 ........................CTTTTATCCTTAGGTGGAGAA 585
                              |||:::|||:::|||
1479 luAsnIleCysLysCysPheHisPheHisSerLeuGlyLeuAsnGlyGly 1495

 586 ACCTATATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAATATT 635
          |||:::::::::|||              ||||||
1496 LeuProGluIleAsnValAsnLys............AspLys..... 1505

 636 GAAGCTTGTTAATGATTTTGATTTAGATGGTGTA.............G 670
      :::|||  |||::::::   :::   :::                :
1506 .AsnLeuTyrAsnAsnLeuHisIleThrAsnCysLeuLeuPheLysLysG 1522

 671 ATATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAAT 720
     ::         |||         ||||||         ||||||:::|||
1522 luThrThrGluGluValLeuLysLysPheLeuProAsnAsnGluIleAsn 1538

 721 TTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCC 770
     :::   :::||||||            |||   :::         ::: |
1539 IleMetSerTyrTyr.........TyrAsnHisIleLeuTyrArgLeu.A 1552

 771 GGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCAT...TAT 817
     ||   |||:::   ::::::   ::::::|||:::   ||||||   |||
1552 rgMetLysAsnLysTyrGluAspIleIleHisAspLysLeuHisValTyr 1568

 818 CATGCGTTTCAGGAGTTG.............................. 835
     |||   ::::::||||||
1569 HisLysLeuLysGluLeuIleLysTyrGlnTyrLysGluTyrLeuLeuHi 1585

 836 ....................................CATCTTTCTGTAA 848
                                         |||::::::   |
1585 sLysThrValTyrProArgAsnIleCysArgAsnGluHisMetAsnGlnL 1602

 849 AG.........ATGAAGAATCTCCATATAACACT................ 873
     ||            |||::::::::::|||:::
1602 ysAspAsnCysThrLysAspIleTyrIleAsnGluAspAsnAsnLysThr 1618

 874 ..............AAATTTTTGTCTGAACAAATAGAAACAAATAAAGA 908
                   |||:::   :::|||:::   |||   ||||||::
1619 GluLeuAsnIleGluLysIleSerLysGluAsnAsnGluGluAsnLysAs 1635

 909 ATTACATAGGGCAGCAGCGATGTTATCA......GCAGGAACTTTTATTA 952
     :   :::      ::::::      :::         |||:::::|||
1635 nThrTyrMetAsnThrThrSerTyrLysGluLeuLeuGlyAsnTyrIleA 1652

 953 ATATTTTTAATACA.....................GCAAAGGAGAAA 978
     ||:::::::::|||                     :::||||||||||
1652 snPheLeuAspThrPheAsnLeuTyrAspAsnIleTyrSerLysGluLys 1668

 979 ATA..............GATCTTGTA...................... 990
                      ||||||:::
1669 TyrGluThrAspGluAsnAspLeuIleLeuAsnAsnLysGluProSerIl 1685

 991 ..........TTTATTCAAACATACAAT.........TTAGAAACTACAA 1021
           |||   ::::::||||||         |||:::::   |
1685 eSerTyrAsnPheAsnSerAsnTyrAsnAsnAspLeuLeuLysSerAspA 1702

1022 AT.................CCAGATATAATGGTAGATATGTACTTA... 1050
     ||                 ||||||      |||   |||
1702 snValTyrGluTyrIleTyrLysAspIleTyrTyrAspSerTyrTyrAsp 1718

1051 ......TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATT 1094
           :::::::::|||:::      ||||||::::::
1719 LysAsnThrTyrIleTyrTyrAspAsnLysTyrThrPhe........... 1731

1095 AGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAG 1144
                    |||:::   ::::::|||      :::::::::
```

```
1732 ..............HisLysThrAsnSerPheIleAsnAspGluAsnG 1743

1145 AATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAAT 1194
               :::|||:::         :::::::|||:::   :::|||
1743 lyCysTyrHisLeuLeuThrTyrProLeuGluAspGluIleGluAsnMet 1759

1195 AAT 1197
     |||
1760 Asn 1760
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:AAW02159

seq_documentation_block:

ID AAW02159 standard; Protein; 866 AA.
XX
AC AAW02159;
XX
DT 14-JAN-1997 (first entry)
XX
DE Soluble chitinase.
XX
KW Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW catabolism.
XX
OS Vibrio furnissii.
XX
PN WO9625424-A1.
XX
PD 22-AUG-1996.
XX
PF 13-FEB-1996; 96WO-US02332.
XX
PR 13-FEB-1995; 95US-0386727.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Bassler B, Chitlaru E, Keyhani N, Roseman S, Rowe C;
PI Yu C;
XX
DR WPI; 1996-393335/39.
DR N-PSDB; AAT36390.
XX
PT Chitin biosynthetic enzymes end I, exo I and exo II - are
PT periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT aryl beta-N-acetyl:gluco:amidase(s), respectively
XX
PS Example 4; Page 79-82; 101pp; English.
XX
CC Periplasmic chitodextrinase (AAW02156), periplasmic
CC Beta-N-acetylglucosaminidase (AAW02157) and aryl
CC Beta-N-acetylglucosaminidase (AAW02158) can be used to produce chitin
CC oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC higher, by contacting them with soluble chitin. The enzymes are
CC encoded by the genes endI, exoI and exoII respectively. They are
CC all genes involved in the catabolic pathway of chitin.
XX
SQ Sequence 866 AA;

alignment_scores:

```
            Quality:  130.00            Length:    359
              Ratio:  0.812               Gaps:     13
 Percent Similarity:  44.568   Percent Identity: 18.942
```

alignment_block:

US-09-579-383-2 x AAW02159 ..

Align seg 1/1 to: AAW02159 from: 1 to: 866

```
 211 GATGGGTGT..................................... 219
     |||||||||
 118 AspGlyCysThrLeuSerAspLysLysGluIleValValAlaAspThrAs 134
```

```
220 .AACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGATAGAA 268
134 pGlySerHisLeuAlaProLeuAsnAlaProLeuGlnGluAsnAsnLys. 150

269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATT 318
151 .....ProTyrThrAsnLysAlaGlyLysVal 159

319 ATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACA 368
160 ValGlyAlaTyrTyrValGluTrpGlyValTyrGly......ArgLysPh 174

369 TATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCT. 417
174 eThrValAspLysIleProAlaLysAsnLeuThrHisIleLeuTyrGlyP 191

418 .....CGCATTAAT...ATGTTATATGATGTA 441
191 heThrProIleCysGlyGlyAsnGlyIleAsnAspSerLeuLysGluIle 207

442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
208 SerGlySerPheGluAlaLeuGlnArgSerCysAlaGlyArgGluAspPh 224

492 AGAATAT..... 498
224 eLysValSerIleHisAspProTrpAlaAlaValGlnMetGlyGlnGlyA 241

499 .....GAAACCTATGGTATGATGCTTAATGAAATTAGA 531
241 snLeuThrAlaPheAspGluProTyrLysGlyAsnPheGlyAsnLeuMet 257

532 CGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGG 581
258 AlaLeuLysLysAlaAsnProAsnLeuLysIleLeuProSerValGlyGl 274

582 AGAAACC.....TATATGATAGATATAGAAAAAGAAA 613
274 yTrpThrLeuSerAspProPheTyrPhePheSerAspLysThrLysArgA 291

614 TTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGAT...TTA 660
291 spThrPheValAlaSerMetLysGluTyrLeuGlnThrTrpLysPhePhe 307

661 GATGGTGTAGATATTGACTGGGAA.....CCACATGGGAAGTTTTA 701
308 AspGlyValAspIleAspTrpGluPheProGlyGlyGlnGlyAlaAsnPr 324

702 CAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACT 751
324 oAsnLeuGlyGlyProAsnAspGlyAlaThrTyrValAlaLeuMetLysG 341

752 TGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCA 801
341 luLeuArgAlaMetLeuAspGlu..... 348

802 TCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGA 851

348 ..... 348

852 TGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAA 901
349 .....LeuGluAlaGluThrG 354

902 ATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATT 951
354 lyArgGlnTyrGluLeuThrSerAlaIleSerAlaGlyGlyAspLysIle 370

952 AAT......ATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTAT 995
371 AlaLysValAspTyrGlnAlaAlaGlnGlnTyrMetAspTyrIlePheLe 387

996 TCAAACATACAAT.....TTAGAAACTACAAATCCAGATATAA 1033
387 uMetSerTyrAspPheSerGlyAlaPheAspLeuLysAsnLeuAlaHisG 404

1034 TGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATAT..... 1077
404 lnThrAsnLeuTyrAlaSerSerTrpAspProAlaThrLysTyrThrThr 420

1078 ......AACATCACAATCATATTAGGT 1098
421 AspLysGlyValLysAlaLeuLeuGly 429
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52307

seq_documentation_block:

ID AAY52307 standard; Protein; 866 AA.
XX
AC AAY52307;
XX
DT 09-FEB-2000 (first entry)
XX
DE Vibrio furnissii extracellular chitinase.
XX
KW Chitinase; extracellular; endI; Endo-I; chitin; chitosan;
KW beta 1-4 N-acetylglucosamine; GlcNAc; degradation; catabolism;
KW oligosaccharide; agriculture; medicine; nitrogen fixing nodules;
KW disease resistance; induction; fungicide; antimetastatic;
KW Lewis lung carcinoma; immune system; macrophage; activation; production;
KW recombinant protein.
XX
OS Vibrio furnissii.
XX
FH Key Location/Qualifiers
FT Peptide 1..24
FT Protein 25..866
FT /note= "Mature extracellular chitinase"
XX
PN US5985644-A.
XX
PD 16-NOV-1999.
XX
PF 13-FEB-1996; 96US-0600452.
XX
PR 13-FEB-1995; 95US-0386727.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Bassler B, Chitlaru E, Yu C, Roseman S, Keyhani NO;
XX
DR WPI; 2000-022280/02.
DR N-PSDB; AAZ38243.
XX
PT DNA encoding periplasmic chitodextrinase endoenzyme -
XX
PS Example 5; Columns 51-56; 37pp; English.
XX
CC This sequence represents Vibrio furnissii extracellular chitinase,
CC which is encoded by the chiA gene. Chitin is a homopolymer of beta 1-4
CC N-acetylglucosamine (GlcNAc). Chitin degradation by V. furnissii
CC involves several signal transducing systems and a multitude of proteins,
CC in contrast to other organisms, which only require two enzymes to
CC degrade chitin to GlcNAc. Chitinase degrades chitin into GlcNAc dimers.
CC Chitin oligosaccharides have been recently shown to be physiologically
CC active and are useful in agriculture and medicine. Derivatised
CC oligosaccharides are generated by Rhizobium species as signals
CC for the formation of nitrogen fixing nodules by leguminous plants
CC and also induce disease resistance in certain plants. They also inhibit
CC the growth of several fungal pathogens. The GlcNAc hexamer is a potent
CC antimetastatic agent against Lewis lung carcinoma, and GlcNAc polymers of
CC varying lengths activate macrophages and the immune system. Prior art
CC methods of production of such oligosaccharides are prohibitively
CC expensive and there are limitations in resolving mixtures of these
CC compounds. Use of recombinant V. furnissii chitin catabolic enzymes

```
CC   may help to reduce or eliminate these problems.
XX
SQ   Sequence   866 AA;


alignment_scores:
            Quality:  130.00              Length:    359
              Ratio:  0.812                Gaps:     13
 Percent Similarity:  44.568   Percent Identity: 18.942

alignment_block:
 US-09-579-383-2 x AAY52307   ..

 Align seg 1/1  to: AAY52307  from: 1  to: 866

     211 GATGGGTGT..................................... 219
         ||||||||
     118 AspGlyCysThrLeuSerAspLysLysGluIleValValAlaAspThrAs 134

     220 .AACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGATAGAA 268
          :::|||:::::::::      |||      :::::::::|||::::::
     134 pGlySerHisLeuAlaProLeuAsnAlaProLeuGlnGluAsnAsnLys. 150

     269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATT 318
                                 |||   ::::::         :::
     151 .....................ProTyrThrAsnLysAlaGlyLysVal 159

     319 ATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACA 368
         :::::::::||||||   :::|||:::      |||          |||
     160 ValGlyAlaTyrTyrValGluTrpGlyValTyrGly......ArgLysPh 174

     369 TATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCT. 417
           :::|||:::   |||      ::::::   :::|||   ::::::
     174 eThrValAspLysIleProAlaLysAsnLeuThrHisIleLeuTyrGlyP 191

     418 .....................CGCATTAAT...ATGTTATATGATGTA 441
                                 ||||||      |||   ::::::
     191 heThrProIleCysGlyGlyAsnGlyIleAsnAspSerLeuLysGluIle 207

     442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
         |||      |||::::::   ||||||            ::::::   ::
     208 SerGlySerPheGluAlaLeuGlnArgSerCysAlaGlyArgGluAspPh 224

     492 AGAATAT..................................... 498
         ::::
     224 eLysValSerIleHisAspProTrpAlaAlaValGlnMetGlyGlnGlyA 241

     499 ................GAAACCTATGGTATGATGCTTAATGAAATTAGA 531
                         |||   |||         ::::::::::::
     241 snLeuThrAlaPheAspGluProTyrLysGlyAsnPheGlyAsnLeuMet 257

     532 CGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGG 581
           ::::::|||::::::|||::::::   ||||||   |||:::|||||
     258 AlaLeuLysLysAlaAsnProAsnLeuLysIleLeuProSerValGlyGl 274

     582 AGAAACC.................TATATGATAGATATAGAAAAAGAAA 613
         |   |||                 ::::::   |||      |||:::
     274 yTrpThrLeuSerAspProPheTyrPhePheSerAspLysThrLysArgA 291

     614 TTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGAT...TTA 660
           :::|||   ::::::   :::   ::::::   :::       :::
     291 spThrPheValAlaSerMetLysGluTyrLeuGlnThrTrpLysPhePhe 307

     661 GATGGTGTAGATATTGACTGGGAA.........CCACATGGGAAGTTTTA 701
         ||||||||||||||||||||||||            :::|||
     308 AspGlyValAspIleAspTrpGluPheProGlyGlyGlnGlyAlaAsnPr 324

     702 CAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACT 751
          ||||||:::       |||   :::       |||:::   |||::::::
     324 oAsnLeuGlyGlyProAsnAspGlyAlaThrTyrValAlaLeuMetLysG 341

     752 TGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCA 801
           ||||||      :::   |||

     341 luLeuArgAlaMetLeuAspGlu..................... 348

     802 TCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGA 851

     348 ..................................................... 348

     852 TGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAA 901
                                                 :::   ||||||:
     349 .............................LeuGluAlaGluThrG 354

     902 ATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATT 951
         :::::::::   :::   ::::::|||:::   ::::::|||      |||
     354 lyArgGlnTyrGluLeuThrSerAlaIleSerAlaGlyGlyAspLysIle 370

     952 AAT......ATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTAT 995
                   ::::::::::|||::::::   :::|||   :::|||::
     371 AlaLysValAspTyrGlnAlaAlaGlnGlnTyrMetAspTyrIlePheLe 387

     996 TCAAACATACAAT............TTAGAAACTACAAATCCAGATATAA 1033
         :::::::|||:::           ::::::      |||         :
     387 uMetSerTyrAspPheSerGlyAlaPheAspLeuLysAsnLeuAlaHisG 404

    1034 TGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATAT...... 1077
         ::::::::::::|||   |||            :::   ||||||
     404 lnThrAsnLeuTyrAlaSerSerTrpAspProAlaThrLysTyrThrThr 420

    1078 ......AACATCACAATCATATTAGGT 1098
                ::::::      :::||||||
     421 AspLysGlyValLysAlaLeuLeuGly 429

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY52533

seq_documentation_block:
ID   AAY52533 standard; Protein; 509 AA.
XX
AC   AAY52533;
XX
DT   22-FEB-2000  (first entry)
XX
DE   D. pteronyssius 98 kD mite allergen protein (map) PDerp98-509.
XX
KW   Mite allergen protein; map; high molecular weight; HMW-map; allergy;
KW   house dust mite; IgE; immunoglobulin E; allergen; mapB;
KW   hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
KW   canine; veterinary; antibody; vaccine; immunisation.
XX
OS   Dermatophagoides pteronyssius.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "Signal peptide"
FT   Protein         20..509
FT                   /note= "Mature PDerp98-509"
XX
PN   WO9954349-A2.
XX
PD   28-OCT-1999.
XX
PF   16-APR-1999;   99WO-US08524.
XX
PR   17-APR-1998;   98US-0062013.
PR   13-MAY-1998;   98US-0085295.
PR   02-SEP-1998;   98US-0098909.
XX
PA   (HESK-) HESKA CORP.
XX
PI   McCall CA,  Hunter SW,  Weber ER;
XX
DR   WPI; 2000-052700/04.
DR   N-PSDB; AAZ38585, AAZ38586, AAZ38587, AAZ38588.
XX
PT   Novel high molecular weight Dermatophagoides nucleic acid polypeptides
PT   used to modify an animals' hypersensitivy to mite allergens -
```

XX
PS Claim 3; Page 134-136; 154pp; English.
XX
CC This sequence represents Dermatophagoides pteronyssius mite allergen
CC protein (map) PDerp98-509. PDerp98-509 has a molecular weight of 98 kD,
CC comprising 509 amino acids, and has a high degree of homology with the
CC D. farinae 98 kD allergen, mapB (AAY52523). Nucleic acid molecules
CC encoding PDerp98-509 were isolated from a D. pteronyssius cDNA library
CC by hybridisation with a probe encoding the D. farinae high molecular
CC weight map (HMW-map) composition. Mite allergenic proteins and
CC peptides, and nucleic acids encoding them, may be used in therapeutic
CC compositions to modify an animal's hypersensitivity reaction to mite
CC allergens. Animals that may be treated include mammals and birds,
CC especially felines, canines, equines, humans, other pets, and work or
CC domestic animals. The proteins or fragments may also be used to
CC diagnose allergies via a skin test. The proteins and peptides can also
CC be used to raise antibodies, which have a variety of potential
CC uses. For example, they can be used as vaccines to passively immunise
CC animals against dust mite hypersensitivity, as positive controls in
CC test kits and as tools to recover desired dust mite allergens from a
CC mixture of proteins.
XX
SQ Sequence 509 AA;

alignment_scores:
Quality: 126.50 Length: 519
Ratio: 0.583 Gaps: 25
Percent Similarity: 41.811 Percent Identity: 18.304

alignment_block:
US-09-579-383-2 x AAY52533 ..

Align seg 1/1 to: AAY52533 from: 1 to: 509

```
193 TATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAA 242
    ::::::   :::|||      :::|||   :::   ::::::      ::
  5 PheAlaLeuPheCysIleTrpAlaCysIleGlyLeuMetAsnAlaAlaTh 21

243 TAAAAATATAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGG 292
    :|||:::   :::::::|||      |||:::|||   :::|||:::
 21 rLysArgAspHisAsnAsnTyrSerLysAsnProMetArgIleVal.... 36

293 AGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGG 342
                                      |||   |||:::|||
 37 ..............................CysTyrValGlyThrTrp 42

343 AACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGT 392
    :::               :::   |||   :::|||         :::
 43 Ser..............ValTyrHisLysValAspProTyrThrIleGl 54

393 GTCAATT.....................TTATATATTGCATTTGCTCGCA 421
      :::|||                  |||      :::|||||:::|
 54 uAspIleAspProPheLysCysThrHisLeuMetTyrGlyPheAlaLysI 71

422 TTAAT.........ATGTTATATGATGTATCTAGACCATTTAATGGAAGA 462
    ||:::                  :::|||      ||||||:::
 71 leAspGluTyrLysTyrThrIleGlnValPheAspProPheGlnAspAsp 87

463 CAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGTAT 512
    ::::::         :::|||||||||   |||
 88 AsnHisAsnSerTrpGluLysHisGlyTyrGlu................. 98

513 GATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTA 562
               ::::::::::::::|||      :::|||:::::::
 99 .........ArgPheAsnAsnLeuArgLeuLysAsnProGluLeuThrT 112

563 TTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGAA 612
      ::::::||||||||||||      :::            ||||||
112 hrMetIleSerLeuGlyGly......TrpTyrGluGlySerGluLysTyr 126

613 ATTGAT.............................TATGTGGATAAAAT 632
       |||                            :::|||::::::::
```

```
127 SerAspMetAlaAlaAsnProThrTyrArgGlnGlnPheValGlnSerVa 143

633 ATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGG 682
    :|||   ::::::::::::::::   :::||||||:::|||:::|||||||
143 lLeuAspPheLeuGlnGluTyrLysPheAspGlyLeuAspLeuAspTrpG 160

683 AACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTAT 732
    ||      |||:::      :::      :::::::::   :::||||||
160 luTyrProGlySerArgLeuGlyAsnProLysIleAspLysGlnAsnTyr 176

733 TATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTT 782
                ::::::|||:::|||:::
177 ............LeuThrLeuValArgGlu..................... 182

783 AATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAG 832

182 ..................................................... 182

833 TTGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTG 882
                  |||:::      :::|||::::::
183 ...........LeuLysGluAlaPheGluProPheGly............ 191

883 TCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTT 932
                                      :::      :::|||   ::
192 ............................TyrLeuLeuThrAlaAlaVa 198

933 ATCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAG 982
    :|||   |||                                |||:::|||||||
198 lSerProGly.........................LysAspLysIleA 206

983 ATCTTGTATTT..................................... 993
    ||:::::::::
206 spValAlaTyrGluLeuLysGluLeuAsnGlnLeuPheAspTrpMetAsn 222

994 ATTCAAACATACAATTTAGAA............................ 1014
    ::::::||||||:::   :::
223 ValMetThrTyrAspTyrHisGlyGlyTrpGluAsnValPheGlyHisAs 239

1015 .........ACTACAAATCCAGATATAATGGTAGATATGTACTTATCCC 1054
                  :::||||||         ::::::         |
239 nAlaProLeuTyrLysArgProAspGluThrAspGluLeu.........H 253

1055 ATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCA 1104
     ||   ||||||:::::::::|||::::::        |||
253 isThrTyrPheAsnValAsnTyrThrMetHisTyrTyrLeu......... 266

1105 TTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTAGA 1154
           |||:::|||:::         :::             :::
267 .........AsnAsnGlyAlaThrArgAspLysLeuValMetGlyValPr 280

1155 ATTGGTAGGAAAA......ACAATACATGATAAAAATCAAAATAATAATA 1198
     :::   |||:::      :::|||:::|||:::::: |||:::   :::
280 oPheTyrGlyArgAlaTrpSerIleGluAspArgSer.LysValLysLeu 296

1199 GGGCAGATGGTATAGGGATATGGCA.........TTTATTTAT...GAAA 1236
     |||:::   :::   ||| |||:::         ||||||      |||:
297 GlyAspProAlaLysGly.MetSerProProGlyPheIleThrGlyGluG 313

1237 GAACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAAATAT 1286
     ::   :::   ::::::|||      |||::: :::|||      :::
313 luGlyValLeuSerTyrIleGluLeuCysGln.LeuPheGlnLysGluGl 329

1287 TTGGAAACATTTAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAA 1336
      |||   |||::::::
329 uTrp...HisIleGln.................................. 333

1337 CTGAAAACCCTGAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTC 1386
                                        |||||||||
334 ........................TyrAspGluTyr............ 337

1387 GTTATTCCAACCATAGGGATATATTACAAACACAATGATGCTATATGGAA 1436
                    :::      |||   :::||||||   ||||||
338 ............TyrAsnAlaProTyrGlyTyrAsnAspLysIleTrp.. 349
```

```
1437 AACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGG 1486

 349 ..................................................  349

1487 ACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCAT 1536
               |||   |||:::   :::::::::   ::::::
 350 ...........ValGlyTyrAspAspLeuAlaSerIleSerCysLysLeu  362

1537 TATTATAACACTGACTATAAAGAAAGCTCTATTATTATATGGAAAGGGGA 1586
       :::      :::         |||::::::::::||||||:::   ||
 363 AlaPheLeuLysGluLeuGlyValSerGlyValMetIleTrpSerLeuGl  379

1587 A 1587
     |
 379 u 379
```

OM of: US-09-579-383-2 to: Issued_Patents_AA:* out_format : pfs

Date: Mar 21, 2002 4:40 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09579383/runat_20032002_151239_29300/app_query.fasta_1.1849
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09579383_@CGN1_1_26 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-579-383-2
Query length: 1764
Database: Issued_Patents_AA:*
Database sequences: 212252
Database length: 22503292
Search time (sec): 68.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation ..		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-358-901-2 +			142.50	225.95	6.9e-06	561	!	S	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-566-347-2 +			142.50	225.95	6.9e-06	561	!	S	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-693-835-2 +			142.50	225.95	6.9e-06	561	!	S	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-386-727-8 +			130.00	198.50	0.0002	866	!	S	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-600-452A-8 +			130.00	198.50	0.0002	866	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-929-329-6 +			126.50	191.34	0.0004	934	!	S	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-855A-2 +			118.00	165.21	0.0039	2391	!		
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-150-741-2 +			118.00	165.21	0.0039	2391	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-480-604A-10 +			114.50	159.05	0.0087	2366	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-405-496A-10 +			114.50	159.05	0.0087	2366	!		
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-915-136-10 +			114.50	159.05	0.0087	2366	!		
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-308-375-2 +			113.50	157.66	0.0108	2285	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-949-588-2 +			112.50	169.39	0.0077	711	!	S	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-912-4 +			109.00	151.97	0.0275	1861	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-569-853-1 +			109.00	149.86	0.0300	2233	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-987-439-1 +			109.00	149.86	0.0300	2233	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-912-3 +			108.50	150.45	0.0317	1964	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-912-2 +			108.50	149.94	0.0323	2052	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-416-581B-1 +			108.00	160.39	0.0226	770	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-416-581B-5 +			108.00	160.39	0.0226	770	!		
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-087-465-6 +			108.00	160.39	0.0226	770	!	S	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-276-099A-14 +			108.00	160.38	0.0226	771	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-781-890-14 +			108.00	160.38	0.0226	771	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-323-170B-2 +			107.50	143.24	0.0500	3135	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-480-604A-6 +			107.00	144.03	0.0523	2710	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-405-496A-6 +			107.00	144.03	0.0523	2710	!		
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-915-136-6 +			107.00	144.03	0.0523	2710	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-568-459A-4 +			106.50	150.49	0.0431	1435	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-826B-4 +			106.50	150.49	0.0431	1435	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-052-778-15 +			106.00	162.98	0.0276	452	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-369-796-12 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-416-581B-9 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-852-091-12 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-820-754-12 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-956-652-12 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-956-869-12 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-012-710-8 +			106.00	156.81	0.0357	770	!	S	
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-948-547-12 +			106.00	156.81	0.0357	770	!		
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-364-970-3 +			106.00	156.81	0.0357	770	!	S	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-364-970-5 +			106.00	156.81	0.0357	770	!	S	
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-17025-12 +			106.00	156.81	0.0357	7			
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-569-853-2 +			106.00	144.48	0.0599	2233			
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-307-499-15 -			103.00	157.91	0.0543	440			
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-299-268-15 -			103.00	157.91	0.0543	440			
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-198-446B-15 +			103.00	138.42	0.1229	2368			

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-358-901-2

seq_documentation_block:
; Sequence 2, Application US/08358901
; Patent No. 5554521
; GENERAL INFORMATION:
; APPLICANT: Suslow, Trevor V.
; APPLICANT: Jones, Jonathan D.G.
; TITLE OF INVENTION: No. 5554521el Chitinase-Producing Plants
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,901
; FILING DATE: 19-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/550,253
; FILING DATE: 09-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/888,033
; FILING DATE: 18-JUL-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/593,691
; FILING DATE: 26-MAR-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 12176-5-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-358-901-2

alignment_scores:
Quality: 142.50 Length: 425
Ratio: 0.720 Gaps: 20
Percent Similarity: 46.588 Percent Identity: 20.471

alignment_block:
US-09-579-383-2 x US-08-358-901-2 ..

Align seg 1/1 to: US-08-358-901-2 from: 1 to: 561

```
      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :
```

```
 27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
    :::::      |||              ::::::::::::      :::
 44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCCGGAGAT.. 213
    |||:::      :::      |||:::::::::   :::   ||||||
 57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
     |||   :::   :::|||      |||:::|||
 73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

247 ..............................AATATAAATAAAAATGAT 264
                                   ::::::||||||:::
 90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

265 AGA.......................................... 267
    |||
107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

267 .......................................... 267

123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
         |||:::|||      :::|||||||:::   |||      ||||||
140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
    :::          ||||||||||||:::   :::|||:::      ||
155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
    |         :::      :::|||:::   |||      ::::::   :
171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
    :::::   ::::::   |||               |||   :::   |||
186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

439 ..............................GTATCTAGACCATTTAA 455
                                  ::::::|||      ::
203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
    :||||||:::   |||   :::      |||            :::|||:
219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

494 AATATGAAACCTATGGTATGATG......................... 516
    ::           |||:::
236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
    ::::::::::::      ::::::::::::::::||||||:::   |||||
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
    |   |||:::||||||   |||                  :::|||   |
269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
    ||      |||:::      :::|||   ::::::   ::::::::::::
286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
       :::      :::||||||||||||||||||||||||   |||   ||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
    ||||            ||||||::::::   :::   ::::::   |||:
319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
    ::   |||::::::   ||||||      :::   :::
336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                  ||||||:::   :::|||
349 ..........................LeuSerAlaGluThrGly..... 354

837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                              |||:::   |||:
355 ................................ArgLysTyrGluLeuT 360

884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
    :::::   |||:::::::::|||:::            :::|||
360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                 :::|||:::|||:::::::::::||
375 ........................TyrAsnValAlaGlnAsnSerMetAs 383

984 TCTTGTATTTATTCAAACATACAAT 1008
    |   :::|||:::::::::|||:::
383 pHisIlePheLeuMetSerTyrAsp 391
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-566-347-2

```
seq_documentation_block:
; Sequence 2, Application US/08566347
; Patent No. 5633450
;  GENERAL INFORMATION:
;    APPLICANT:  Suslow, Trevor V.
;    APPLICANT:  Jones, Jonathan D.G.
;    TITLE OF INVENTION:  No. 5633450el Chitinase-Producing Plants
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/566,347
;      FILING DATE:  01-DEC-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/358,901
;      FILING DATE:  19-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/550,253
;      FILING DATE:  09-JUL-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/888,033
;      FILING DATE:  18-JUL-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/593,691
;      FILING DATE:  26-MAR-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, William M.
;      REGISTRATION NUMBER:  30,223
;      REFERENCE/DOCKET NUMBER:  12176-5-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 amino acids
```

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-566-347-2


alignment_scores:
            Quality:  142.50              Length:    425
              Ratio:  0.720                Gaps:     20
 Percent Similarity:  46.588   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x US-08-566-347-2   ..

 Align seg 1/1  to: US-08-566-347-2  from: 1  to: 561

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::     |||              :::::::::::::     :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   :::   :::|||      |||:::|||
      73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ..............................AATATAAATAAAAATGAT 264
                                      ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA........................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 ............................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
              |||:::|||      :::||||||:::   |||      ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::         ::::::::::::|||:::   :::|||:::      ||
     155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         :::      :::|||:::   |||    ::::::   :
     171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
         :::::   ::::::   |||                |||   :::   |||
     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

     439 ..............................GTATCTAGACCATTTAA 455
                                      :::::::|||      ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
         :||||||:::   |||   :::      |||            :::|||:
     219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

     494 AATATGAAACCTATGGTATGATG........................... 516
```

```
         ::           |||:::
     236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::::::|||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
         |   |||:::||||||   |||                    :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
         :::      :::||||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

     693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
         ||||            ||||||::::::   :::   ::::::   |||:
     319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

     737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
         ::   |||::::::   ||||||      :::   :::
     336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                       ||||||::   :::|||
     349 ........................LeuSerAlaGluThrGly..... 354

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                              |||:::   |||:
     355 ................................ArgLysTyrGluLeuT 360

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         :::::   |||:::::::::|||:::            :::|||
     360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                    :::|||:::|||:::::::::::||
     375 ......................TyrAsnValAlaGlnAsnSerMetAs 383

     984 TCTTGTATTTATTCAAACATACAAT 1008
         |   :::|||:::::::::|||:::
     383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-693-835-2

seq_documentation_block:
; Sequence 2, Application US/08693835
; Patent No. 5776448
;  GENERAL INFORMATION:
;    APPLICANT:  Suslow, Trevor V.
;    APPLICANT:  Jones, Jonathan D.G.
;    TITLE OF INVENTION:  No. 5776448el Chitinase-Producing Plants
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/693,835
;      FILING DATE:  01-AUG-1996
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/358,901
;      FILING DATE:  19-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/550,253
;      FILING DATE:  09-JUL-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/888,033
;      FILING DATE:  18-JUL-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/593,691
;      FILING DATE:  26-MAR-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, William M.
;      REGISTRATION NUMBER:  30,223
;      REFERENCE/DOCKET NUMBER:  12176-5-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-693-835-2


alignment_scores:
            Quality:  142.50              Length:    425
              Ratio:  0.720                Gaps:     20
 Percent Similarity:  46.588   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x US-08-693-835-2   ..

 Align seg 1/1  to: US-08-693-835-2  from: 1  to: 561

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::       ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||              ::::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||:::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   :::   ::::||      |||:::|||
      73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ..............................AATATAAATAAAAATGAT 264
                                        ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA........................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 ............................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
              |||:::|||      :::|||||||:::   |||      ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::         ::::::::::::|||:::   :::|||:::      ||
     155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         :::      :::|||:::   |||      ::::::   :
     171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
         :::::   ::::::   |||              |||   :::   |||
     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

     439 ................................GTATCTAGACCATTTAA 455
                                        :::::::|||     ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
         :|||||:::   |||   :::      |||           :::|||:
     219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

     494 AATATGAAACCTATGGTATGATG........................... 516
         ::          |||:::
     236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::::|||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC..................TATATGATAG 598
         |   |||:::||||||   |||                  :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
            :::      :::|||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

     693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
         ||||            ||||||::::::   :::   ::::::   |||:
     319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

     737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
         ::   |||::::::   ||||||      :::   :::
     336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                       |||||||::   :::|||
     349 ............................LeuSerAlaGluThrGly..... 354

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                            |||:::   |||:
     355 ..................................ArgLysTyrGluLeuT 360

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         :::::   |||::::::::::|||:::            :::|||
     360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                    :::|||:::|||:::::::::::||
     375 ........................TyrAsnValAlaGlnAsnSerMetAs 383

     984 TCTTGTATTTATTCAAACATACAAT 1008
         |   :::|||:::::::::|||:::
     383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-386-727-8

seq_documentation_block:
; Sequence 8, Application US/08386727
```

```
; Patent No. 5792647
;  GENERAL INFORMATION:
;    APPLICANT:  ROSEMAN, SAUL
;    APPLICANT:  BASSLER, BONNIE
;    APPLICANT:  KEYHANI, NEMAT O.
;    APPLICANT:  CHITLARU, EDITH
;    APPLICANT:  ROWE, CHRIS
;    APPLICANT:  YU, CHARLES
;    TITLE OF INVENTION:  BACTERIAL CATABOLISM OF CHITIN
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/386,727
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HOBBS, ANN S.
;      REGISTRATION NUMBER:  36,830
;      REFERENCE/DOCKET NUMBER:  4130/206916
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  866 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-386-727-8


alignment_scores:
            Quality:  130.00              Length:    359
              Ratio:  0.812                Gaps:     13
 Percent Similarity:  44.568   Percent Identity: 18.942

alignment_block:
 US-09-579-383-2 x US-08-386-727-8   ..

 Align seg 1/1  to: US-08-386-727-8  from: 1  to: 866

     211 GATGGGTGT..................................... 219
         |||||||||
     118 AspGlyCysThrLeuSerAspLysLysGluIleValValAlaAspThrAs 134

     220 .AACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGATAGAA 268
          :::|||:::::::::      |||      ::::::::::|||::::::
     134 pGlySerHisLeuAlaProLeuAsnAlaProLeuGlnGluAsnAsnLys. 150

     269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATT 318
                                  |||   ::::::         :::
     151 ......................ProTyrThrAsnLysAlaGlyLysVal 159

     319 ATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACA 368
         :::::::::||||||   :::|||:::      |||       |||
     160 ValGlyAlaTyrTyrValGluTrpGlyValTyrGly......ArgLysPh 174

     369 TATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCT. 417
             :::|||:::   |||      ::::::   :::|||   ::::::
     174 eThrValAspLysIleProAlaLysAsnLeuThrHisIleLeuTyrGlyP 191

     418 ......................CGCATTAAT...ATGTTATATGATGTA 441
                                  ||||||      |||   ::::::
     191 heThrProIleCysGlyGlyAsnGlyIleAsnAspSerLeuLysGluIle 207

     442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
         |||      |||::::::   ||||||            ::::::   ::
     208 SerGlySerPheGluAlaLeuGlnArgSerCysAlaGlyArgGluAspPh 224

     492 AGAATAT..................................... 498
         ::::
     224 eLysValSerIleHisAspProTrpAlaAlaValGlnMetGlyGlnGlyA 241

     499 ................GAAACCTATGGTATGATGCTTAATGAAATTAGA 531
                         |||   |||          ::::::::::::
     241 snLeuThrAlaPheAspGluProTyrLysGlyAsnPheGlyAsnLeuMet 257

     532 CGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGG 581
           ::::::|||:::::::|||::::::   ||||||   |||:::|||||
     258 AlaLeuLysLysAlaAsnProAsnLeuLysIleLeuProSerValGlyGl 274

     582 AGAAACC.................TATATGATAGATATAGAAAAAGAAA 613
         |   |||                 ::::::   |||      |||:::
     274 yTrpThrLeuSerAspProPheTyrPhePheSerAspLysThrLysArgA 291

     614 TTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGAT...TTA 660
              :::|||   ::::::   :::   ::::::   :::      :::
     291 spThrPheValAlaSerMetLysGluTyrLeuGlnThrTrpLysPhePhe 307

     661 GATGGTGTAGATATTGACTGGGAA.........CCACATGGGAAGTTTTA 701
         |||||||||||||||||||||||||           :::|||
     308 AspGlyValAspIleAspTrpGluPheProGlyGlyGlnGlyAlaAsnPr 324

     702 CAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACT 751
          ||||||:::      |||   :::     |||:::   |||::::::
     324 oAsnLeuGlyGlyProAsnAspGlyAlaThrTyrValAlaLeuMetLysG 341

     752 TGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCA 801
           ||||||      :::   |||
     341 luLeuArgAlaMetLeuAspGlu..................... 348

     802 TCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGA 851

     348 ..................................................... 348

     852 TGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAA 901
                                              :::   ||||||:
     349 ..........................LeuGluAlaGluThrG 354

     902 ATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATT 951
         ::::::::   :::   :::::::|||:::   :::::::|||      |||
     354 lyArgGlnTyrGluLeuThrSerAlaIleSerAlaGlyGlyAspLysIle 370

     952 AAT......ATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTAT 995
                 :::::::::|||::::::   :::|||   :::|||::
     371 AlaLysValAspTyrGlnAlaAlaGlnGlnTyrMetAspTyrIlePheLe 387

     996 TCAAACATACAAT............TTAGAAACTACAAATCCAGATATAA 1033
         :::::::|||:::           ::::::      |||        :
     387 uMetSerTyrAspPheSerGlyAlaPheAspLeuLysAsnLeuAlaHisG 404

    1034 TGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATAT...... 1077
         ::::::::::|||   |||            :::   ||||||
     404 lnThrAsnLeuTyrAlaSerSerTrpAspProAlaThrLysTyrThrThr 420

    1078 ......AACATCACAATCATATTAGGT 1098
                ::::::      :::||||||
     421 AspLysGlyValLysAlaLeuLeuGly 429

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-600-452A-8

seq_documentation_block:
; Sequence 8, Application US/08600452A
```

```
; Patent No. 5985644
;  GENERAL INFORMATION:
;    APPLICANT:  ROSEMAN, SAUL
;    APPLICANT:  BASSLER, BONNIE
;    APPLICANT:  KEYHANI, NEMAT O.
;    APPLICANT:  CHITLARU, EDITH
;    APPLICANT:  ROWE, CHRIS
;    APPLICANT:  YU, CHARLES
;    TITLE OF INVENTION:  BACTERIAL CATABOLISM OF CHITIN
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/600,452A
;      FILING DATE:  13-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07662/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  866 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-600-452A-8


alignment_scores:
            Quality:  130.00              Length:    359
              Ratio:  0.812                Gaps:     13
 Percent Similarity:  44.568   Percent Identity: 18.942

alignment_block:
 US-09-579-383-2 x US-08-600-452A-8   ..

 Align seq 1/1  to: US-08-600-452A-8  from: 1  to: 866

     211 GATGGGTGT..................................... 219
         |||||||||
     118 AspGlyCysThrLeuSerAspLysLysGluIleValValAlaAspThrAs 134

     220 .AACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGATAGAA 268
          :::|||:::::::::      |||      :::::::::|||::::::
     134 pGlySerHisLeuAlaProLeuAsnAlaProLeuGlnGluAsnAsnLys. 150

     269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATT 318
                                  |||   ::::::         :::
     151 ......................ProTyrThrAsnLysAlaGlyLysVal 159

     319 ATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACA 368
         :::::::::||||||   :::|||:::     |||        |||
     160 ValGlyAlaTyrTyrValGluTrpGlyValTyrGly......ArgLysPh 174

     369 TATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCT. 417
             :::|||:::   |||      ::::::   :::|||   ::::::
     174 eThrValAspLysIleProAlaLysAsnLeuThrHisIleLeuTyrGlyP 191

     418 ......................CGCATTAAT...ATGTTATATGATGTA 441
                                      ||||||      |||   ::::::
     191 heThrProIleCysGlyGlyAsnGlyIleAsnAspSerLeuLysGluIle 207

     442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
         |||      |||::::::   ||||||             ::::::   ::
     208 SerGlySerPheGluAlaLeuGlnArgSerCysAlaGlyArgGluAspPh 224

     492 AGAATAT.......................................... 498
         ::::
     224 eLysValSerIleHisAspProTrpAlaAlaValGlnMetGlyGlnGlyA 241

     499 .................GAAACCTATGGTATGATGCTTAATGAAATTAGA 531
                         |||   |||         ::::::::::::
     241 snLeuThrAlaPheAspGluProTyrLysGlyAsnPheGlyAsnLeuMet 257

     532 CGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGG 581
           ::::::|||::::::|||::::::   ||||||   |||:::|||||
     258 AlaLeuLysLysAlaAsnProAsnLeuLysIleLeuProSerValGlyGl 274

     582 AGAAACC.................TATATGATAGATATAGAAAAAGAAA 613
         |   |||                 ::::::   |||      |||:::
     274 yTrpThrLeuSerAspProPheTyrPhePheSerAspLysThrLysArgA 291

     614 TTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGAT...TTA 660
             :::|||   ::::::   :::   ::::::   :::       :::
     291 spThrPheValAlaSerMetLysGluTyrLeuGlnThrTrpLysPhePhe 307

     661 GATGGTGTAGATATTGACTGGGAA.........CCACATGGGAAGTTTTA 701
         ||||||||||||||||||||||||            :::|||
     308 AspGlyValAspIleAspTrpGluPheProGlyGlyGlnGlyAlaAsnPr 324

     702 CAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACT 751
          ||||||:::      |||   :::      |||:::   |||::::::
     324 oAsnLeuGlyGlyProAsnAspGlyAlaThrTyrValAlaLeuMetLysG 341

     752 TGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCA 801
           ||||||      :::   |||
     341 luLeuArgAlaMetLeuAspGlu........................... 348

     802 TCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGA 851

     348 ................................................... 348

     852 TGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAA 901
                                                  :::   ||||||:
     349 ...............................LeuGluAlaGluThrG 354

     902 ATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATT 951
         ::::::::   :::   ::::::|||:::   :::::::|||      |||
     354 lyArgGlnTyrGluLeuThrSerAlaIleSerAlaGlyGlyAspLysIle 370

     952 AAT......ATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTAT 995
                   :::::::::|||::::::   :::|||   :::|||::
     371 AlaLysValAspTyrGlnAlaAlaGlnGlnTyrMetAspTyrIlePheLe 387

     996 TCAAACATACAAT............TTAGAAACTACAAATCCAGATATAA 1033
         :::::::|||:::           ::::::      |||       :
     387 uMetSerTyrAspPheSerGlyAlaPheAspLeuLysAsnLeuAlaHisG 404

    1034 TGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATAT...... 1077
         ::::::::::|||   |||            :::   ||||||
     404 lnThrAsnLeuTyrAlaSerSerTrpAspProAlaThrLysTyrThrThr 420

    1078 ......AACATCACAATCATATTAGGT 1098
                ::::::      :::||||||
     421 AspLysGlyValLysAlaLeuLeuGly 429

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-929-329-6

seq_documentation_block:
; Sequence 6, Application US/08929329
```

```
; Patent No. 6120770
;  GENERAL INFORMATION:
;    APPLICANT:  Adams, John H
;    APPLICANT:  Dalton, John P
;    APPLICANT:  Kappe, Stefan
;    TITLE OF INVENTION:  Plasmodium Proteins Useful for Preparing
;    TITLE OF INVENTION:  Vaccine Compositions
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Barnes & Thornburg
;      STREET:  11 S Meridian
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/929,329
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Breen, John P
;      REGISTRATION NUMBER:  38,833
;      REFERENCE/DOCKET NUMBER:  835910-28685
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 231-7745
;      TELEFAX:  (317) 231-7433
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  934 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium berghei
US-08-929-329-6


alignment_scores:
            Quality:  126.50              Length:    504
              Ratio:  0.591                Gaps:     19
 Percent Similarity:  42.460   Percent Identity: 18.651

alignment_block:
 US-09-579-383-2 x US-08-929-329-6   ..

 Align seg 1/1  to: US-08-929-329-6  from: 1  to: 934

      70 AAAGGAAAAAATAATATAAATAATTCATTGGGAATA.........ATACG 110
         :::::::|||   |||      |||:::||||||::          :::::
     394 GlnAsnLysLeuAsnSerMetAsnAspLeuGlyValCysValLeuLeuLy 410

     111 GGAAAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCAC 160
         :|||:::   :::   :::   :::   :::   :::   ||||||::::
     410 sGluArgProThrCysValLeuLysLysGlnAsnTyrTyrSerPheThrA 427

     161 ATCTTAAATCGAAT.........AATAGTAATTTTGTAGAATATGGATCT 201
         ::|||   :::|||          |||::::::   :::||||||   :::
     427 snLeuThrAlaAsnTyrPheAspAsnAsnGlnAsnIleGluTyrProAsp 443

     202 TAT............TGCGGAGATGGGTGTAACTCTAGAATTACAAAAAA 239
                         |||   :::   ||||||::::::::   ||
     444 IleGluAsnValLysIleGlyLysAsnArgAsnSerGluLeuSerGlyAs 460

     240 T.........................AATAAAAATATAAATAAAAATG 262

         |                          |||:::||||||||||||:::
     460 nLeuThrTyrAsnAspLysLysPheLysAsnSerAsnIleAsnLysGlyM 477

     263 ATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
             |||   ::::::|||   |||      ::::::   |||
     477 etValIleSerMetAsnAspIleSerGluIleLysGluAsnSerLys... 492

     313 GGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGC 362
                               :::   :::::::::|||   :::::
     493 ..............LeuGlnThrAsnLysArgAsnGluGlyLysLysTh 504

     363 AAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCAT 412
         :|||:::   :::   :::   |||::::::   |||   |||:::
     504 rLysTyrGlyLeuTyrAsnTyrProIleThrProIleSerTyrLeu.... 519

     413 TTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGAAGA 462
           :::||||||         ::::::::::::   :::
     520 .....GlnIleAsnHisLysArgGluLeuThrLysLysTyrMetTyrPhe 534

     463 CAAAGATTCCTATTAAGAAAACAC............GGCTTAGAATATGA 500
         ::::::   |||      :::|||            :::::::::|||||
     535 GluAsnSerLeuThrProSerHisAsnThrGlyAlaSerIleArgTyrGl 551

     501 A.........ACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAA 541
         |              |||:::      |||:::   ||||||
     551 uGlyAsnAsnArgLeuGlyIleAspAlaAsnAsnAsnArgArgAsnThrT 568

     542 AAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTAT 591
                 |||:::   :::       :::   :::   :::
     568 yrGlyThrGlnAspIleAsnLeuLysArgAsnAsnAsnTyrLysGlnPro 584

     592 ATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCT 641
                 ::::::   :::   :::|||:::|||::::::
     585 LysAsnLysLeuAsnProGlnAlaGluTyrMetAspArgPhe........ 598

     642 TGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATG 691
           |||::::::   :::   :::   |||||||||:::      |
     599 .......AspIleGluLysAsnHisIleTyrIleAspTrpLysGlnAspG 613

     692 GGAAG..................................................... 696
         |||||
     613 lyLysTyrGlySerAspLysLeuLysTyrAsnIleIleSerHisGluThr 629

     696 ..................................................... 696

     630 AlaAsnThrValGlnSerLeuLeuIleThrAspLysAsnAspIleCysPr 646

     696 ..................................................... 696

     646 oAsnHisTyrSerProGlyArgAlaGlnGlySerCysProAsnTyrGlyL 663

     697 ..........................................TTTTAC 702
                                                  ::::::
     663 ysSerIleIleValLysAlaLeuGluGlyThrAsnGlyAspGluTyrPhe 679

     703 AACTTAAATGAATTAAAT..............TTTTCAAATTATTATAT 737
         |||||||||   ||||||                ::   |||   |||::
     680 AsnLeuAsnPheLeuAsnGluIleArgThrGlyTyrLeuAsnArgTyrMe 696

     738 TAAATTA..................................... 744
         :|||
     696 tLysTyrAspValGluLeuProTyrGluLysSerGlyLeuAlaMetHisH 713

     745 ........ATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
             :::|||          |||::::::   ||||||:::|||
     713 isGlyAspLeuAsnGluCysProLysSerLeuAspGluGluAsnLeuTyr 729

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
         :::|||::::::::::   |||   ::::::
     730 LysIleLysSerAspTyrAsnTyrGlyMet..................... 739

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTGTCTG 886
             ||||||:::               |||:::   |||   :::
```

```
     740 .......CysLysSerThrValLeuLysSerAsnValProPheLysThrT 754

     887 AACAAATAGAAACAAATAAAGAATTACAT..................... 915
           :::   :::|||:::|||   |||:::
     754 yrAsnTyrArgThrLysLysLeuLeuTyrPheGlyLeuTyrGlyLeuGly 770

     916 ...AGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAA 962
           |||   ::::::   :::|||        :::   |||||||||::
     771 GlyArgLeuGlySerAsnMetSer......LysIleLysAsnIlePheLy 785

     963 TACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATACAATTTAG 1012
         ::::      :::::::|||   |||   ::::::
     785 sSerArgProAsnAsnIleThrLeuProMetPhe.............. 796

    1013 AAACTACAAATCCAGATATAATGGTAGATATGTACTTATCCCATTTATAT 1062
                |||||||:::::::::   ::::::           ||||||
     797 ........AsnProSerLeuIleLysAsnLeuLeuAspCysSerLeuTyr 810

    1063 TTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAACA 1112
                              :::      ||||||     ||||||::
     811 ....................SerTyrCysLeuGlyProCysLeuGluAs 820

    1113 TAACAGAGGTGGA.........................TTTAGTCCCG 1135
         :      ::::::                          |||:::   |
     820 nAlaTyrAsnAsnLysCysPheArgAsnLeuProAlaTyrPheAsnHisG 837

    1136 AAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAAT 1185
         ||:::||||||   :::   |||              |||:::::::::
     837 luThrLysGluCysValIleLeu.........GlyThrHisGluGlnGlu 850

    1186 CAAAATAATAAT 1197
         :::   |||:::
     851 ArgValAsnAsp 854
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-446-855A-2

```
seq_documentation_block:
; Sequence 2, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29.009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2391 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-446-855A-2


alignment_scores:
            Quality:  118.00            Length:    457
              Ratio:  0.554               Gaps:     22
 Percent Similarity:  46.608   Percent Identity: 20.569

alignment_block:
 US-09-579-383-2 x US-08-446-855A-2   ..

 Align seg 1/1  to: US-08-446-855A-2  from: 1  to: 2391

      46 TATTCTGCAAATTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTC 95
         ||||||:::::   ::      ||||||:::|||:::|||:::
      42 TyrSerValGlyTyrGluGlyCysLysGlyAsnAsnSerIleSer..... 56

      96 ATTGGGAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATAC 145
                                                        |
      57 ..............................................CysH 58

     146 ATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATAT 195
         ||:::::::::   ::::::   :::|||:::::::::      :::
      58 isLysGluTyrArgAsnIleIleAsnAsnAspAsnSerLysAsnSerAsn 74

     196 GGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAA 245
         :::|||:::|||                              :::|||:::::
      75 AsnSerPheCys.........................AsnAsnGluGl 82

     246 AAATATAAATAAAAATGAT......AGAAAATCACCAAGACAAATTTTAG 289
         :|||      |||:::|||        |||:::   |||        |
      82 uAsnAsnLeuLysAspAspLeuLeuTyrLysAsnSerArgLeuGluAsnG 99

     290 AGGAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCA 339
         ||::::::                 |||:::::::|||
      99 luAspPhe.................IleValThrGlyGluValIlePhe 109

     340 TGGAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAAT 389
            :::::::::           :::  :::   |||   :::   ::
     110 AsnThrAlaMetValGlyTyrProGluAlaLeuThrAspProSerTyrPh 126

     390 GGTGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATG 439
         :   :::||||||   ::::::|||      |||:::      |||   :
     126 eGlyGlnIleLeuValLeuThrPheProSerIleGlyAsn...TyrGlyI 142

     440 TATCTAGA..............CCATTTAATGGAAGACAAAGATTCCTA 474
         ::::::::             |||:::      |||:::|||
     142 leGluLysValLysHisAspGluThrPheGlyLeuValGlnAsnPheGlu 158

     475 TTAAGAAAA.........CACGGCTTA...................... 492
            :::|||         :::||||||
     159 SerAsnLysIleGlnValGlnGlyLeuValIleCysGluTyrSerLysGl 175

     493 .......GAATATGAAACCTATGGTATGATGCTTAATGAAATTAGACGTA 535
               :::|||::::::|||   :::   |||:::|||      :::|
     175 nSerTyrHisTyrAsnSerTyr...IleThrLeuSerGluTrpLeuLysI 191

     536 TCAGAAAAGTA.......................CGTCCAGATGTAATT 561
         ||   |||:::                      |||      :::
     191 leTyrLysIleProCysIleGlyGlyIleAspThrArgAlaLeuThrLys 207

     562 ATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGA 611
         :::|||   :::   |||:::            |||   |||   |||::
     208 LeuLeuArgGluLysGlySerMetLeuGlyLysIleValIleTyrLysAs 224

     612 AATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAG 661
         :   :::::::::::|||:::   |||   :::|||   ||||||
     224 nArgGlnHisIleAsnLysLeuTyrLysGluIleAsnLeuPheAspProG 241

     662 ATGGTGTAGAT..................................... 672
```

```
      :::::::|||
   241 lyAsnIleAspThrLeuLysTyrValCysAsnHisPheIleArgValIle 257

   673 ............ATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAA 710
                    |||   :::            ::::::   |||   ::
   258 LysLeuAsnAsnIleThrTyr............AsnTyrLysAsnLysGl 270

   711 TGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAA 760
       :|||:::|||:::::::|||      |||                  :::
   270 uGluPheAsnTyrThrAsnGluMetIle..............ThrAsnA 282

   761 AAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCT 810
         :::      |||:::         :::|||:::||||||   :::
   282 spSerSerMetGluAspHisAspAsnGluIleAsnGlySerIleSerAsn 298

   811 GCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAA...GA 857
             :::|||   |||   ::::::|||||   |||:::|||   ::
   299 PheAsnAsnCysProSer...IleSerSerPheAspLysSerGluSerLy 314

   858 ATCTCCATATAACACTAAATTTTTGTCTGAACAAATA..............  894
       ::::      |||      ::::||   :::::::::
   314 sAsnValIleAsnHisThrLeuLeuArgAspLysMetAsnLeuIleThrS 331

   895 ........GAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCA 936
               |||      |||:::||||||::::::
   331 erSerGluGluTyrLeuLysAspLeuHisAsnCysAsn............. 343

   937 GCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCT 986
                        |||:::::::::::::::|||   |||
   344 ....................PheSerAsnSerSerAspLysAsnAsp.. 352

   987 TGTATTTATTCAAACATACAATTTAGAAACTACAAATCCAGATATAATGG 1036
           |||::::::   |||::::::
   353 .SerPhePheLysLeuTyrGlyIle..................CysGluT 363

  1037 TAGATATGTACTTATCCCATTTA.......................... 1059
         |||   ||||||      |||
   363 yrAspLysTyrLeuIleAspLeuGluGluAsnAlaSerPheHisTyrAsn 379

  1060 ..................TATTTTGGTTTAAAATATAACATCACAATCAT 1091
                         |||:::   ::::::   |||   :::|||::
   380 AsnValAspGluTyrGlyTyrTyrAspValAsnLysAsnThrAsnIleLe 396

  1092 ATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATA 1141
       :   :::   :::::::|||:::|||:::::::::         ||||||:
   396 uSerAsnAsnLysIleGluGlnAsnAsnAsnAsn.........GluAsnA 410

  1142 AAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAAT 1191
       :::::                          ::::::||||||:::|||
   410 snLys..........................AsnAsnLysAsnAsnAsn 417

  1192 AATAATAGGGCAGATGGTATA 1212
       ||||||::::::|||   |||
   418 AsnAsnGluValAspTyrIle 424

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-150-741-2

seq_documentation_block:
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;  LENGTH: 2391
;  TYPE: PRT
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2


alignment_scores:
           Quality:  118.00          Length:    457
             Ratio:  0.554             Gaps:     22
 Percent Similarity:  46.608   Percent Identity: 20.569

alignment_block:
 US-09-579-383-2 x US-09-150-741-2   ..

 Align seg 1/1  to: US-09-150-741-2  from: 1  to: 2391

     46 TATTCTGCAAATTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTC 95
        ||||||::::::   :::      ||||||:::|||:::|||:::
     42 TyrSerValGlyTyrGluGlyCysLysGlyAsnAsnSerIleSer..... 56

     96 ATTGGGAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATAC 145
                                                       . |
     57 .............................................CysH 58

    146 ATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATAT 195
        ||::::::::::   ::::::   :::|||::::::::::      :::
     58 isLysGluTyrArgAsnIleIleAsnAsnAspAsnSerLysAsnSerAsn 74

    196 GGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAA 245
        :::|||:::|||                              :::|||:::::
     75 AsnSerPheCys..........................AsnAsnGluGl 82

    246 AAATATAAATAAAAATGAT......AGAAAATCACCAAGACAAATTTTAG 289
        :|||      |||:::|||         |||:::   |||         |
     82 uAsnAsnLeuLysAspAspLeuLeuTyrLysAsnSerArgLeuGluAsnG 99

    290 AGGAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCA 339
        ||::::::                  |||:::::::|||
     99 luAspPhe.................IleValThrGlyGluValIlePhe 109

    340 TGGAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAAT 389
           :::::::::            :::   :::   |||   :::   ::
    110 AsnThrAlaMetValGlyTyrProGluAlaLeuThrAspProSerTyrPh 126

    390 GGTGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATG 439
        :   :::||||||   :::::::|||      |||:::      |||   :
    126 eGlyGlnIleLeuValLeuThrPheProSerIleGlyAsn...TyrGlyI 142

    440 TATCTAGA..............CCATTTAATGGAAGACAAAGATTCCTA 474
        ::::::::                  |||:::      |||:::|||
    142 leGluLysValLysHisAspGluThrPheGlyLeuValGlnAsnPheGlu 158

    475 TTAAGAAAA.........CACGGCTTA...................... 492
           :::|||         :::||||||
    159 SerAsnLysIleGlnValGlnGlyLeuValIleCysGluTyrSerLysGl 175

    493 .......GAATATGAAACCTATGGTATGATGCTTAATGAAATTAGACGTA 535
              :::|||::::::|||   :::   |||:::|||      :::|
    175 nSerTyrHisTyrAsnSerTyr...IleThrLeuSerGluTrpLeuLysI 191

    536 TCAGAAAAGTA.......................CGTCCAGATGTAATT 561
        ||   |||:::                       |||      :::
    191 leTyrLysIleProCysIleGlyGlyIleAspThrArgAlaLeuThrLys 207

    562 ATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGA 611
        :::|||   :::   |||:::            |||   |||   |||::
    208 LeuLeuArgGluLysGlySerMetLeuGlyLysIleValIleTyrLysAs 224
```

```
     612 AATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAG 661
         :   :::::::::::::|||:::   |||   :::|||   ||||||
     224 nArgGlnHisIleAsnLysLeuTyrLysGluIleAsnLeuPheAspProG 241

     662 ATGGTGTAGAT..................................... 672
           ::::::|||
     241 lyAsnIleAspThrLeuLysTyrValCysAsnHisPheIleArgValIle 257

     673 ............ATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAA 710
                     |||   :::          ::::::   |||   ::
     258 LysLeuAsnAsnIleThrTyr............AsnTyrLysAsnLysGl 270

     711 TGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAA 760
         :|||:::|||:::::::|||      |||                 :::
     270 uGluPheAsnTyrThrAsnGluMetIle..............ThrAsnA 282

     761 AAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCT 810
           :::      |||:::         :::|||:::||||||   :::
     282 spSerSerMetGluAspHisAspAsnGluIleAsnGlySerIleSerAsn 298

     811 GCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAA...GA 857
               :::|||   |||   ::::::||||||   |||:::|||   ::
     299 PheAsnAsnCysProSer...IleSerSerPheAspLysSerGluSerLy 314

     858 ATCTCCATATAACACTAAATTTTTGTCTGAACAAATA.............. 894
         ::::      |||      :::|||   :::::::::
     314 sAsnValIleAsnHisThrLeuLeuArgAspLysMetAsnLeuIleThrS 331

     895 ........GAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCA 936
                 |||      |||:::||||||:::::
     331 erSerGluGluTyrLeuLysAspLeuHisAsnCysAsn............ 343

     937 GCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCT 986
                           |||:::::::::::::::|||   |||
     344 .....................PheSerAsnSerSerAspLysAsnAsp.. 352

     987 TGTATTTATTCAAACATACAATTTAGAAACTACAAATCCAGATATAATGG 1036
             |||::::::   |||::::::   '
     353 .SerPhePheLysLeuTyrGlyIle..................CysGluT 363

    1037 TAGATATGTACTTATCCCATTTA.......................... 1059
           |||   ||||||      |||
     363 yrAspLysTyrLeuIleAspLeuGluGluAsnAlaSerPheHisTyrAsn 379

    1060 ..................TATTTTGGTTTAAAATATAACATCACAATCAT 1091
                           |||:::   ::::::   |||   :::|||::
     380 AsnValAspGluTyrGlyTyrTyrAspValAsnLysAsnThrAsnIleLe 396

    1092 ATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATA 1141
         :   :::   ::::::|||:::|||:::::::::          ||||||:
     396 uSerAsnAsnLysIleGluGlnAsnAsnAsnAsn.........GluAsnA 410

    1142 AAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAAT 1191
         :::::                              ::::::||||||:::|||
     410 snLys..........................AsnAsnLysAsnAsnAsn 417

    1192 AATAATAGGGCAGATGGTATA 1212
         ||||||::::::|||   |||
     418 AsnAsnGluValAspTyrIle 424

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-480-604A-10

seq_documentation_block:
; Sequence 10, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2366 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-10


alignment_scores:
            Quality:  114.50            Length:    654
              Ratio:  0.400               Gaps:     40
 Percent Similarity:  43.731   Percent Identity: 20.183

alignment_block:
 US-09-579-383-2 x US-08-480-604A-10   ..

 Align seg 1/1  to: US-08-480-604A-10  from: 1  to: 2366

      33 AGTATCCATCTT.................GTATTCTGCAAATTCCAGAA 64
         |||:::||||||                :::   ||||||:::||
    1554 SerValHisLeuAspGluSerGlyValAlaGluIleLeuLysPheMetAs 1570

      65 CCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGAA 114
         |    ||||||::::::|||   ::::::||||||      :::   |||
    1570 n.ArgLysGlyAsnThrAsnThrSerAspSerLeuMetSerPheLeuGlu 1586

     115 AATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCT 164
         :::   |||                  |||   :::   |||   :::::
    1587 SerMetAsn.................IleLysSerIlePheValAsnPh 1597
```

```
 165 TAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGATG 214
1597 eLeuGlnSerAsnIleLysPheIleLeuAspAlaAsnPheIleIleSerG 1614

 215 GGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGAT 264
1614 lyThrThrSer...IleGlyGlnPheGluPheIleCysAspGluAsnAsp 1629

 265 AGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGG 314
1630 AsnIleGlnPro..................................... 1633

 315 TATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAA 364
1634 .............TyrPheIleLysPheAsnThr................ 1640

 365 AACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTT 414
1641 ........LeuGluThrAsn.........TyrThrLeuTyrValGly... 1650

 415 GCTCGCATTAATATGTTA............TATGATGTA........... 441
1651 AsnArgGlnAsnMetIleValGluProAsnTyrAspLeuAspAspSerGl 1667

 442 ..........TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAA 481
1667 yAspIleSerSerThrValIleAsnPheSerGlnLysTyrLeu....... 1681

 482 AACACGGCTTAGAA................................TAT 498
1682 ..TyrGlyIleAspSerCysValAsnLysValValIleSerProAsnIle 1697

 499 GAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACG 548
1698 TyrThrAspGluIleAsnIleThrProValTyrGluThrAsnAsnThrTy 1714

 549 TCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAG 598
1714 rProGluValIleValLeu............................... 1720

 599 ATATAGAAAAAGAAATTGATTATGTG...GATAAAATATTGAAGCTTGTT 645
1721 ...........AspAlaAsnTyrIleAsnGluLysIleAsnValAsnIle 1733

 646 AATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAA 695
1734 AsnAspLeuSerIleArgTyrVal.........TrpSerAsnAspGlyAs 1747

 696 GTTTTACAACTTA............AATGAATTAAATTTTTCAAATTATT 733
1747 nAspPheIleLeuMetSerThrSerGluGluAsnLysValSerGlnValL 1764

 734 ATATTAAATTAATTAACTTGTTAAGA...AAAACTATTCCGGAAGAAAAG 780
1764 ysIleArgPheValAsnValPheLysAspLysThrLeuAlaAsnLysLeu 1780

 781 TTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCA......TGCGT 824
1781 SerPheAsnPheSerAspLysGlnAspValProValSerGluIleIleLe 1797

 825 TTCAGGAGTTGCATCTTTCTGTAAAGATGAA................... 855
1797 uSerPheThrProSerTyrTyrGluAspGlyLeuIleGlyTyrAspLeuG 1814

 856 .....GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 900
1814 lyLeuValSerLeuTyrAsnGluLysPheTyr.................. 1824

 901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTAT 950
1825 .........IleAsnAsnPheGlyMetMetValSerGlyLeuIleTyrIl 1838

 951 TAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
1838 eAsn.........................AspSerLeuTyr....... 1843

1001 CATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050
1843 ..................................................... 1843

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
1844 .........TyrPheLysProProValAsn...AsnLeuIleThrGlyPh 1856

1101 T...TCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAAT 1147
1856 eValThrValGlyAspAspLysTyrTyrPheAsnProIleAsnGlyGlyA 1873

1148 TATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAAT 1197
1873 laAlaSerIleGlyGluThrIleIleAspAspLysAsnTyrTyrPheAsn 1889

1198 AGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACC 1247
1890 GlnSer.........GlyVal.......................LeuGl 1895

1248 AACTGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGGAAACATT 1297
1895 nThrGlyValPheSerThrGlu............AspGlyPheLysTyrP 1908

1298 TAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCT 1347
1908 heAlaPro..................................... 1910

1348 GAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCCAAC 1397
1911 ......AlaAsnThrLeuAspGluAsnLeuGluGlyGluAlaIleAspPh 1925

1398 CATAGGG.....................ATATATTAC.........AAAC 1417
1925 eThrGlyLysLeuIleIleAspGluAsnIleTyrTyrPheAspAspAsnT 1942

1418 ACAATGATGCTATA...TGGAAA........................... 1437
1942 yrArgGlyAlaValGluTrpLysGluLeuAspGlyGluMetHisTyrPhe 1958

1438 .....................................ACTAG 1442
1959 SerProGluThrGlyLysAlaPheLysGlyLeuAsnGlnIleGlyAspTy 1975

1443 ATCTTATTCAATTCATGCACCTGGTGTA...................... 1470
1975 rLysTyrTyrPheAsnSerAspGlyValMetGlnLysGlyPheValSerI 1992

1471 ........GACAGATATGAATGGGAC.........TTGGTCAAAGTATGC 1503
1992 leAsnAspAsnLysHisTyrPheAspAspSerGlyValMetLysValGly 2008

1504 TATGAAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGACTA 1553
2009 TyrThrGluIle...AspGlyLys......HisPheTyrPheAlaGluAs 2022

1554 TAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTAAAT 1603
2022 nGlyGluMetGlnIleGlyValPheAsnThrGluAspGly...PheLysT 2038

1604 GGTGGCAACAA...............GGACCTCCGGAAGGTCAGGCACTA 1638
2038 yrPheAlaHisHisAsnGluAspLeuGlyAsnGluGluGlyGluGluIle 2054

1639 GAGTCATACACAAAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATG 1688
2055 ...SerTyrSer.......................GlyIleLeuAsnPh 2062

1689 GAATAAAAAA 1698
```

```
        :|||:::|||
    2062 eAsnAsnLys 2065

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-405-496A-10

seq_documentation_block:
; Sequence 10, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2366 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-10


alignment_scores:
            Quality:  114.50              Length:    654
              Ratio:  0.400                Gaps:     40
 Percent Similarity:  43.731   Percent Identity: 20.183

alignment_block:
 US-09-579-383-2 x US-08-405-496A-10   ..

 Align seg 1/1  to: US-08-405-496A-10  from: 1  to: 2366

      33 AGTATCCATCTT.................GTATTCTGCAAATTCCAGAA 64
         |||:::|||||||                    :::   |||||||:::||
    1554 SerValHisLeuAspGluSerGlyValAlaGluIleLeuLysPheMetAs 1570

      65 CCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGAA 114
         |    ||||||::::::|||   ::::::||||||      :::   |||

1570 n.ArgLysGlyAsnThrAsnThrSerAspSerLeuMetSerPheLeuGlu 1586

 115 AATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCT 164
     :::   |||                  |||   :::   |||   :::::
1587 SerMetAsn.................IleLysSerIlePheValAsnPh 1597

 165 TAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGATG 214
     :   :::::::|||   :::|||:::      :::::::::      :::|
1597 eLeuGlnSerAsnIleLysPheIleLeuAspAlaAsnPheIleIleSerG 1614

 215 GGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGAT 264
     ||   :::|||   |||   :::   :::          ::::::||||||
1614 lyThrThrSer...IleGlyGlnPheGluPheIleCysAspGluAsnAsp 1629

 265 AGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGG 314
     :::   :::|||
1630 AsnIleGlnPro..................................... 1633

 315 TATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAA 364
                 |||:::   :::::|||:::
1634 .............TyrPheIleLysPheAsnThr.............. 1640

 365 AACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTT 414
             ::::::::::|||               ||||||:::::
1641 ........LeuGluThrAsn.........TyrThrLeuTyrValGly... 1650

 415 GCTCGCATTAATATGTTA............TATGATGTA........... 441
        |||   ||||||:::             ||||||:::
1651 AsnArgGlnAsnMetIleValGluProAsnTyrAspLeuAspAspSerGl 1667

 442 ..........TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAA 481
                |||      :::|||      |||:::::|||
1667 yAspIleSerSerThrValIleAsnPheSerGlnLysTyrLeu....... 1681

 482 AACACGGCTTAGAA................................TAT 498
      :::|||::::::
1682 ..TyrGlyIleAspSerCysValAsnLysValValIleSerProAsnIle 1697

 499 GAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACG 548
        |||      :::   ::::::   :::   :::   :::::::::
1698 TyrThrAspGluIleAsnIleThrProValTyrGluThrAsnAsnThrTy 1714

 549 TCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAG 598
      |||:::||||||:::|||
1714 rProGluValIleValLeu............................... 1720

 599 ATATAGAAAAAGAAATTGATTATGTG...GATAAAATATTGAAGCTTGTT 645
                 :::   :::|||:::   :::||||||         :::
1721 ...........AspAlaAsnTyrIleAsnGluLysIleAsnValAsnIle 1733

 646 AATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAA 695
     ||||||:::::::::      |||         |||:::      |||::
1734 AsnAspLeuSerIleArgTyrVal.........TrpSerAsnAspGlyAs 1747

 696 GTTTTACAACTTA............AATGAATTAAATTTTTCAAATTATT 733
     :   :::   |||              :::|||   :::   |||:::
1747 nAspPheIleLeuMetSerThrSerGluGluAsnLysValSerGlnValL 1764

 734 ATATTAAATTAATTAACTTGTTAAGA...AAAACTATTCCGGAAGAAAAG 780
       |||:::::::::|||::::::::::   ||||||:::   ::::::
1764 ysIleArgPheValAsnValPheLysAspLysThrLeuAlaAsnLysLeu 1780

 781 TTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCA......TGCGT 824
        ::::::::::|||   ::::::::::::   :::|||          ::
1781 SerPheAsnPheSerAspLysGlnAspValProValSerGluIleIleLe 1797

 825 TTCAGGAGTTGCATCTTTCTGTAAAGATGAA................... 855
     :|||   :::   |||:::   :::|||
1797 uSerPheThrProSerTyrTyrGluAspGlyLeuIleGlyTyrAspLeuG 1814

 856 .....GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 900
           |||   ||||||   ||||||
1814 lyLeuValSerLeuTyrAsnGluLysPheTyr................... 1824
```

```
 901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTAT 950
1825 .........IleAsnAsnPheGlyMetMetValSerGlyLeuIleTyrIl 1838

 951 TAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
1838 eAsn.........................AspSerLeuTyr....... 1843

1001 CATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050

1843 ................................................. 1843

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
1844 .........TyrPheLysProProValAsn...AsnLeuIleThrGlyPh 1856

1101 T...TCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAAT 1147
1856 eValThrValGlyAspAspLysTyrTyrPheAsnProIleAsnGlyGlyA 1873

1148 TATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAAT 1197
1873 laAlaSerIleGlyGluThrIleIleAspAspLysAsnTyrTyrPheAsn 1889

1198 AGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACC 1247
1890 GlnSer.........GlyVal.......................LeuGl 1895

1248 AACTGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGGAAACATT 1297
1895 nThrGlyValPheSerThrGlu............AspGlyPheLysTyrP 1908

1298 TAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCT 1347
1908 heAlaPro..................................... 1910

1348 GAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCCAAC 1397
1911 ......AlaAsnThrLeuAspGluAsnLeuGluGlyGluAlaIleAspPh 1925

1398 CATAGGG.....................ATATATTAC.........AAAC 1417
1925 eThrGlyLysLeuIleIleAspGluAsnIleTyrTyrPheAspAspAsnT 1942

1418 ACAATGATGCTATA...TGGAAA........................... 1437
1942 yrArgGlyAlaValGluTrpLysGluLeuAspGlyGluMetHisTyrPhe 1958

1438 .............................................ACTAG 1442

1959 SerProGluThrGlyLysAlaPheLysGlyLeuAsnGlnIleGlyAspTy 1975

1443 ATCTTATTCAATTCATGCACCTGGTGTA...................... 1470
1975 rLysTyrTyrPheAsnSerAspGlyValMetGlnLysGlyPheValSerI 1992

1471 ........GACAGATATGAATGGGAC.........TTGGTCAAAGTATGC 1503
1992 leAsnAspAsnLysHisTyrPheAspAspSerGlyValMetLysValGly 2008

1504 TATGAAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGACTA 1553
2009 TyrThrGluIle...AspGlyLys......HisPheTyrPheAlaGluAs 2022

1554 TAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTAAAT 1603
2022 nGlyGluMetGlnIleGlyValPheAsnThrGluAspGly...PheLysT 2038

1604 GGTGGCAACAA..............GGACCTCCGGAAGGTCAGGCACTA 1638
2038 yrPheAlaHisHisAsnGluAspLeuGlyAsnGluGluGlyGluGluIle 2054

1639 GAGTCATACACAAAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATG 1688
2055 ...SerTyrSer.....................GlyIleLeuAsnPh 2062

1689 GAATAAAAAA 1698
2062 eAsnAsnLys 2065
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-915-136-10

seq_documentation_block:

; Sequence 10, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-10

alignment_scores:

```
          Quality:  114.50            Length:    654
            Ratio:  0.400               Gaps:     40
 Percent Similarity:  43.731   Percent Identity: 20.183

alignment_block:
 US-09-579-383-2 x US-08-915-136-10   ..

 Align seg 1/1  to: US-08-915-136-10  from: 1  to: 2366

      33 AGTATCCATCTT.................GTATTCTGCAAATTCCAGAA 64
         |||:::|||||                  :::   ||||||:::||
    1554 SerValHisLeuAspGluSerGlyValAlaGluIleLeuLysPheMetAs 1570

      65 CCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGAA 114
         |    ||||||:::::::|||   ::::::||||||      :::   |||
    1570 n.ArgLysGlyAsnThrAsnThrSerAspSerLeuMetSerPheLeuGlu 1586

     115 AATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCT 164
         :::   |||                  |||   :::   |||   :::::
    1587 SerMetAsn.................IleLysSerIlePheValAsnPh 1597

     165 TAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGATG 214
         :   :::::::|||   :::|||:::      :::::::::      :::|
    1597 eLeuGlnSerAsnIleLysPheIleLeuAspAlaAsnPheIleIleSerG 1614

     215 GGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGAT 264
         ||   :::|||   |||   :::   :::          ::::::||||||
    1614 lyThrThrSer...IleGlyGlnPheGluPheIleCysAspGluAsnAsp 1629

     265 AGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGG 314
         :::   :::|||
    1630 AsnIleGlnPro..................................... 1633

     315 TATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAA 364
                   |||:::   :::::::|||:::
    1634 .............TyrPheIleLysPheAsnThr.............. 1640

     365 AACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTT 414
               ::::::::::|||                |||||||::::::
    1641 ........LeuGluThrAsn.........TyrThrLeuTyrValGly... 1650

     415 GCTCGCATTAATATGTTA............TATGATGTA........... 441
            |||   |||||||:::            |||||||:::
    1651 AsnArgGlnAsnMetIleValGluProAsnTyrAspLeuAspAspSerGl 1667

     442 ..........TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAA 481
                    |||       :::|||       |||::::::|||
    1667 yAspIleSerSerThrValIleAsnPheSerGlnLysTyrLeu....... 1681

     482 AACACGGCTTAGAA................................TAT 498
           :::|||::::::
    1682 ..TyrGlyIleAspSerCysValAsnLysValValIleSerProAsnIle 1697

     499 GAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACG 548
            |||      :::   ::::::   :::   :::   :::::::::
    1698 TyrThrAspGluIleAsnIleThrProValTyrGluThrAsnAsnThrTy 1714

     549 TCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAG 598
          |||:::||||||:::|||
    1714 rProGluValIleValLeu............................... 1720

     599 ATATAGAAAAAGAAATTGATTATGTG...GATAAAATATTGAAGCTTGTT 645
                   :::   :::|||:::   :::||||||          :::
    1721 ...........AspAlaAsnTyrIleAsnGluLysIleAsnValAsnIle 1733

     646 AATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAA 695
         ||||||:::::::::      |||          |||:::      |||::
    1734 AsnAspLeuSerIleArgTyrVal.........TrpSerAsnAspGlyAs 1747

     696 GTTTTACAACTTA............AATGAATTAAATTTTTCAAATTATT 733
         :   :::   |||            :::|||   :::   |||:::
    1747 nAspPheIleLeuMetSerThrSerGluGluAsnLysValSerGlnValL 1764

     734 ATATTAAATTAATTAACTTGTTAAGA...AAAACTATTCCGGAAGAAAAG 780
           |||:::::::::|||:::::::::   ||||||:::   ::::::
    1764 ysIleArgPheValAsnValPheLysAspLysThrLeuAlaAsnLysLeu 1780

     781 TTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCA......TGCGT 824
            ::::::::::|||   ::::::::::::   :::|||         ::
    1781 SerPheAsnPheSerAspLysGlnAspValProValSerGluIleIleLe 1797

     825 TTCAGGAGTTGCATCTTTCTGTAAAGATGAA.................... 855
         :|||   :::   |||:::   :::|||
    1797 uSerPheThrProSerTyrTyrGluAspGlyLeuIleGlyTyrAspLeuG 1814

     856 .....GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 900
                |||   ||||||   ||||||
    1814 lyLeuValSerLeuTyrAsnGluLysPheTyr.................. 1824

     901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTAT 950
                 :::::::::   :::   |||:::|||:::      :::||
    1825 .........IleAsnAsnPheGlyMetMetValSerGlyLeuIleTyrIl 1838

     951 TAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
         ||||                        |||   ::::::
    1838 eAsn.........................AspSerLeuTyr....... 1843

    1001 CATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050

    1843 ..................................................... 1843

    1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
                |||||             |||   :::::::|||   |||||
    1844 .........TyrPheLysProProValAsn...AsnLeuIleThrGlyPh 1856

    1101 T...TCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAAT 1147
         |   ::::::      ::::::      |||:::|||   |||
    1856 eValThrValGlyAspAspLysTyrTyrPheAsnProIleAsnGlyGlyA 1873

    1148 TATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAAT 1197
           ::::::          |||   |||||||||          |||
    1873 laAlaSerIleGlyGluThrIleIleAspAspLysAsnTyrTyrPheAsn 1889

    1198 AGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACC 1247
         ::::::          |||:::                        |||
    1890 GlnSer.........GlyVal.......................LeuGl 1895

    1248 AACTGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGGAAACATT 1297
          ||||||   |||:::::::::              :::   :::|||::::
    1895 nThrGlyValPheSerThrGlu............AspGlyPheLysTyrP 1908

    1298 TAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCT 1347
         ::   |||
    1908 heAlaPro..................................... 1910

    1348 GAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCCAAC 1397
               ::::::|||:::||||||   :::   |||   :::|||
    1911 ......AlaAsnThrLeuAspGluAsnLeuGluGlyGluAlaIleAspPh 1925

    1398 CATAGGG.....................ATATATTAC.........AAAC 1417
            |||                      |||||||||          ::::
    1925 eThrGlyLysLeuIleIleAspGluAsnIleTyrTyrPheAspAspAsnT 1942

    1418 ACAATGATGCTATA...TGGAAA.......................... 1437
         :::::   |||:::   ||||||
    1942 yrArgGlyAlaValGluTrpLysGluLeuAspGlyGluMetHisTyrPhe 1958

    1438 ..........................................ACTAG 1442

    1959 SerProGluThrGlyLysAlaPheLysGlyLeuAsnGlnIleGlyAspTy 1975

    1443 ATCTTATTCAATTCATGCACCTGGTGTA..................... 1470
          :::|||   :::::::::   ||||||
    1975 rLysTyrTyrPheAsnSerAspGlyValMetGlnLysGlyPheValSerI 1992

    1471 ........GACAGATATGAATGGGAC.........TTGGTCAAAGTATGC 1503
```

```
                :::::::::   :::|||         ::::::||||||
    1992 leAsnAspAsnLysHisTyrPheAspAspSerGlyValMetLysValGly 2008

    1504 TATGAAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGACTA 1553
         |||   :::|||   |||||||||      |||:::|||   ::::::
    2009 TyrThrGluIle...AspGlyLys......HisPheTyrPheAlaGluAs 2022

    1554 TAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTAAAT 1603
             |||   :::|||   :::::::::   |||         :::|||:
    2022 nGlyGluMetGlnIleGlyValPheAsnThrGluAspGly...PheLysT 2038

    1604 GGTGGCAACAA..............GGACCTCCGGAAGGTCAGGCACTA 1638
         :::::   :::               |||      ||||||:::   :::
    2038 yrPheAlaHisHisAsnGluAspLeuGlyAsnGluGluGlyGluGluIle 2054

    1639 GAGTCATACACAAAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATG 1688
             ||||||:::                       ||||||   :::::
    2055 ...SerTyrSer.....................GlyIleLeuAsnPh 2062

    1689 GAATAAAAAA 1698
         :|||:::|||
    2062 eAsnAsnLys 2065

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-308-375-2

seq_documentation_block:
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
;  APPLICANT: Genencor International, Inc.
;  TITLE OF INVENTION: Proteases From Gram-Positive Organisms
;  FILE REFERENCE: GC394-PCT
;  CURRENT APPLICATION NUMBER: US/09/308,375
;  CURRENT FILING DATE: 1999-05-14
;  EARLIER APPLICATION NUMBER: EP9719636.4
;  EARLIER FILING DATE: 1997-09-15
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2285
;   TYPE: PRT
;   ORGANISM: Bacillius subtilis
US-09-308-375-2


alignment_scores:
           Quality:  113.50            Length:    474
             Ratio:  0.518               Gaps:     22
 Percent Similarity:  46.203   Percent Identity: 20.675

alignment_block:
 US-09-579-383-2 x US-09-308-375-2   ..

 Align seg 1/1  to: US-09-308-375-2  from: 1  to: 2285

       4 AATTTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTCTGC 53
         |||   |||::::::      ::::::   ::::::      ||||||::
     942 AsnAlaLysValLysAspSerValLeuSerIleAlaAsnAlaTyrSerSe 958

      54 A......AATTCCAGAACCTTGAAAGGAAAA..............AATA 82
         :         |||:::||||||||||   :::               |||:
     958 rIleAspIleSerAsnThrLeuLysThrSerIleSerAspValValAsnL 975

      83 ATATAAAT..................................... 90
         :::::|||
     975 ysLeuAsnLeuLysAspAspLeuAspProGluGluLeuGluLysPheSer 991

      91 AATTCATTGGGAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGA 140
         :::|||||||||   ::::::|||:::   :::|||:::   ::::::
     992 SerSerLeuGlyLysLeuGlnGluLysMetGlnLysAlaLeuAspSerGl 1008

     141 AATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAG 190
             :::::::::|||::::::   |||::::::   ::::::::::::|

    1008 yAspGluLysAlaPheAspAsnAlaLysLysAspLeuGlnSerLeuLeuG 1025

     191 AATATGGATCTTATTGCGGA.........GATGGGTGTAACTCTAGAATT 231
         ||       |||                    |||       :::      :::
    1025 luThrTyrSerLysSerAspSerSerIleAspValPheLysMetSerPhe 1041

     232 ACAAAAAATAATAAAAATATAAATAAAAATGATAGAAAATCACCAAGACA 281
         |||   :::||||||||||:::   :::|||   ||||||      ::
    1042 AspLysAlaGlnLysAsnIleLysAspGlyAsp...LysSerLeuSerSe 1057

     282 AATTTTAGAGGAGTAT............................... 297
         ::::   :::|||
    1057 rValLysSerGluValGlyAspLeuGlyGluThrLeuAlaGluAlaGlyA 1074

     298 ...................AAAAAAAGGAAACAAGGTATTATAGCAGGA 327
                             ||||||   |||:::::::::   |||
    1074 snGluAlaGluAspPheGlyLysLysLeuLysGluAlaLeuAspAla... 1089

     328 TACTATGGTTCATGGAACAGTCAAGGTGAT.........AGAGCAAAACA 368
                         ||||||      |||              |||::
    1090 ..............AsnSerValAspAspIleLysAlaAlaIleLysGl 1101

     369 TATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTC 418
         :|||   |||:::      :::   |||:::      :::      :::
    1101 uMetSerAspAlaMetGlnPheAspSerValGlnAspValLeuAsnGlyA 1118

     419 GCATT...AATATGTTATATGATGTATCTAGACCATTTAATGGAAGACAA 465
           |||   |||          |||          |||:::|||
    1118 spIlePheAsnAsnThrLysAspGlnValAlaProLeuAsnAsp..... 1132

     466 AGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGAT 515
               ||||||:::|||   :::          |||      ::::::
    1133 ......LeuLeuGluLysMetAla.........GluGlyLysSerIleSe 1144

     516 GCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTC 565
             ||||||   :::   :::      ::::::   |||      :::
    1144 rAlaAsnGluAlaAsnThrLeu......IleGlnLysAspLysGluLeuA 1159

     566 TTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGAA... 612
               :::::::::   |||:::   ::::::   |||::::::::::
    1159 laGlnAlaIleSerIleGluAsnGlyValValLysIleAsnArgAspGlu 1175

     613 ..............................................ATTGA 617
                                                      :::
    1176 ValIleLysGlnArgLysValLysLeuAspAlaTyrAsnAspMetValTh 1192

     618 TTATGTGGATAAAATATTGAAG..............CTTGTTAATGATT 652
         |||   :::|||::::::|||                  ::::::   :
    1192 rTyrSerAsnLysLeuMetLysThrGluValAsnAsnAlaIleLysThrL 1209

     653 TTGATTTAGATGGTGTAGATATTGAC............TGGGAACCACAT 690
         :::::   |||   :::.  ||||||              :::   :::
    1209 euAsnAlaAspThrLeuArgIleAspSerLeuLysLysLeuArgLysGlu 1225

     691 GGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAA 740
             |||:::   :::::::::|||   ::::::|||:::      :::||
    1226 ArgLysLeu...AspMetSerGluAlaGluLeuSerAspLeuGluValLy 1241

     741 ATTAATTAACTTGTTA.........AGAAAAACTATTCCG......GAAG 775
         |   ||||||   :::          :::|||   :::          ||||
    1241 sSerIleAsnAsnValAlaAspAlaLysLysGluLeuLysLysLeuGluG 1258

     776 AAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTT 825
         |||||::::::::   :::|||   ||||||:::   ::::::::::::
    1258 luLysMetLeuGlnProGlyGlyTyrSerAsnSerGlnIleGluAlaMet 1274

     826 TCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAA 875
         ::::::|||   |||      ::::::      |||
    1275 GlnSerValLysSerAlaLeuGluSerTyrIleSer.............. 1286

     876 ATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAG 925
             ||||||:::      :::::::::|||::::::::::      |
    1287 ....AlaSerGluGluAlaThrSerThrGlnGluMetAsnLys...GlnA 1301
```

```
     926 CGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAG 975
         ||:::::::::|||||||||   ::::::   ::::::      :::|||
    1301 laLeuValGluAlaGlyThrSerLeuGluAsnTrpThrAspGlnGlnGlu 1317

     976 AAAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAATCC 1025
         |||                          |||   ||||||   :::
    1318 LysAla.....................AsnGluGluThrLysThrSe 1326

    1026 AGATATAATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAAT 1075
                ::::||||||   |||
    1326 rMetTyrValValAspLysTyr..........................  1333

    1076 ATAACATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGA 1125

    1333 ..................................................  1333

    1126 TTTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACA 1175
                         ||||||   ||||||   |||:::      |||
    1334 ..............LysGluAlaLeuGluLysValAsnAlaGluIleAs 1345

    1176 TGATAAAAATCAAAATAATAAT 1197
                ||||::::::   |||
    1345 pLysTyrAsnLysGlnValAsn 1352

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-949-588-2

seq_documentation_block:
; Sequence 2, Application US/08949588
; Patent No. 6025156
;  GENERAL INFORMATION:
;    APPLICANT:  Gwynn, Michael
;    APPLICANT:  Kallendar, Howard
;    APPLICANT:  Palmer, Leslie
;    TITLE OF INVENTION:  No. 6025156el Topoisomerase III
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/949,588
;      FILING DATE:  14-OCT-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/028,417
;      FILING DATE:  15-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50567
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  711 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-949-588-2
```

```
alignment_scores:
           Quality:  112.50            Length:    576
             Ratio:  0.434               Gaps:     31
 Percent Similarity:  44.965   Percent Identity: 20.833

alignment_block:
 US-09-579-383-2 x US-08-949-588-2   ..

 Align seq 1/1  to: US-08-949-588-2  from: 1  to: 711

     250 ATAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAA 299
         :::::::|||      :::|||:::|||   :::   |||         ::
     117 LeuAspLysValGlyAsnLysLysProIleArgArgLeuTrpIleSerSe 133

     300 AAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTC 349
         :      |||:::::::|||      |||:::
     133 rValThrLysLysAlaIleGlnGlnGlyPhe..................  143

     350 AAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATT 399
                       |||:::::::   |||::::::          ::::::
     144 ..............LysAsnLeuLysAspGlyArgGlnTyrAsnAspLeu 155

     400 TTATATATTGCATTTGCTCGC.......................ATTAA 425
            |||   |||:::||||||                        |||||
     156 TyrTyrAlaAlaLeuAlaArgSerGluAlaAspTrpIleValGlyIleAs 172

     426 TATGTTATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTAT 475
         |            ::::::      ::::::::::::           |
     172 nAlaThrArgAlaLeuThrThrLysTyrAspAlaGln............L 185

     476 TAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAA 525
         ||           ::::::      |||      ::::::   ::::::
     185 euSerLeuGlyArgValGlnThrProThrIleGlnLeuValAsnThrArg 201

     526 ATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTT 575
            :::::::|||::::::   :::|||:::      :::   ||||||||
     202 GlnGlnGluIleAsnGlnPheLysProGlnGlnTyrPheThrLeuSerLe 218

     576 AGGTGGAGAAACCTATATGATAGATATAGAAAAAGAAATTGATTATGTGG 625
         |      :::   :::   :::::::::|||::::::      |||::::
     218 uThrValLysGlyPheAspPheGlnLeuGluSerAsnGlnArgTyrThrA 235

     626 ATAAA......ATATTGAAGCTTGTTAATGATTTT...GATTTAGATGGT 666
         ::|||      :::   ::::::||||||::::::   ::::::||||||
     235 snLysGluThrLeuGluGlnMetValAsnAsnLeuLysAsnValAspGly 251

     667 ...GTAGATATTGACTGGGAACCACATGGGAAG..............TT 698
            :::                  |||   |||              ::
     252 LysIleLysSerValAlaThrLysHisLysLysSerTyrProGlnSerLe 268

     699 TTACAACTTAAATGAATTAAAT.........TTTTCAAATTATTATATT. 738
         :|||||||||::::::|||:::         :::   :::|||   |||
     268 uTyrAsnLeuThrAspLeuGlnGlnAspMetTyrArgArgTyrLysIleG 285

     739 ........AAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAG 780
                 :::   :::|||   |||::::::         ::::::|||
     285 lyProLysGluThrLeuAsnThrLeuGlnSerLeuTyrGluArgHisLys 301

     781 TTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGG 830
         :::::::::         :::
     302 ValValThrTyrProArgThr............................  308

     831 AGTTGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTT 880
                                  :::|||   |||   |||
     309 ........................AspSerAsnTyrLeuThr.......  314

     881 TGTCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCA......... 921
           ::::::::::::::::|||   ||||||      :::|||
     315 ..ThrAspMetValAspThrMetLysGluArgIleGlnAlaThrMetAla 330

     922 .................GCAGCGATGTTATCAGCAGGAACTTTT..... 948
                         |||      |||   :::   ||||||
     331 ThrThrTyrLysAspGlnAlaArgProLeuMetSerLysThrPheSerSe 347
```

```
 949 ....ATTAATATTTTTAATACAGCAAAG..................... 972
         ::::::|||||||||:::   |||
 347 rLysMetSerIlePheAsnAsnGlnLysValSerAspHisHisAlaIleI 364

 973 .....GAGAAAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACT 1017
          ::::::      |||:::   :::      |||   |||
 364 leProThrGluValArgProValMetSerAspLeuSerAsnArgGluLeu 380

1018 ACAAATCCAGATATAATGGTAGATATGTACTTATCC.............. 1053
             |||::::::|||:::   :::|||:::
 381 LysLeuTyrAspMetIleValGluArgPheLeuGluAlaLeuMetProPr 397

1054 .CATTTATAT.....................................T 1063
      |||   |||                                    |
 397 oHisGluTyrAspAlaIleThrValThrLeuGluValAlaGlyHisThrP 414

1064 TTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAACAT 1113
     ||   ||||||   |||:::|||   :::|||||||||
 414 heValLeuLysGluAsnValThrThrValLeuGlyPhe............ 426

1114 AACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGG 1163
                              |||::::::   :::      ||
 427 ..........................LysSerIleArgGln......Gl 432

1164 AAAAACAATACATGATAAAAATCAAAATAATAATAGGGCAGATGGTATAG 1213
     |:::::::|||   :::   :::|||      ::::::::::|||
 432 yGluSerIleThrGluMetGlnGlnProPheSerGluGlyAsp....... 446

1214 GGATATGGCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGAT 1263
                                                  :::
 447 ..............................................Glu 447

1264 GTAGATATTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAAGTACA 1313
     |||   |||      |||||||||   :::|||
 448 ValLysIleSerLysThrAsnIleArgGluHis...........GluTh 460

1314 AACTCCAAAAGAC.....................CTTACTATAACTGAAA 1342
      ||||||   :::                    |||        ||||
 460 rThrProProGluTyrPheAsnGluGlySerLeuLeuLysAlaMetGluA 477

1343 ACCCTGAAGACTGTAGCACAATA...GATGAATATGTTCCAGGACTCGTT 1389
     |||||::::::         :::   |||:::   :::   ::::::||
 477 snProGlnAsnPheIleGlnLeuLysAspLys.LysTyrAlaGlnThrLe 493

1390 ATTCCAACCATAGGGATATATTACAAACACAATGATGCTATATGGAAAAC 1439
     |   |||      |||      |||            :::      :::
 493 uLysGlnThrGlyGlyIleGlyThr............ValAlaThrArgA 506

1440 TAGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACT 1489
       |||:::|||:::   :::                 :::|||||||:::
 506 laAspIleIleAspLysLeuPhe............AsnMetAsnAlaIle 518

1490 TGGTCAAAGTATGCT...ATGAAAAAATATGCGATGGGAAAGCAGCCCAT 1536
        |||:::   :::   :::|||      :::   |||||||||   ::
 519 GluSerArgAspGlyLysIleLysValThrSerLysGlyLysGlnIleLe 535

1537 T.............................ATTATAACACTGACTATAA 1556
     :                            ::::::|||        :
 535 uGluLeuAlaProGluGluLeuThrSerProLeuLeuThrAlaGlnTrpG 552

1557 AGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTAAATGGT 1606
     :::::   |||||||||| :::|||||||||     |||      |||   :
 552 luGluLysLeuLeuLeu.IleGluArgGly.LysTyrGlnAlaLysThrP 568

1607 GGCAACAAGGACCTCCGGAAGGTCAGGCACTAGAGTCATACACAAAACTA 1656
     ::                      :::   ::::::::::::||||||
 568 heIle.................AsnGluMetLysAspPheThrLys... 577

1657 GATGCATCCAAATGTCCAGGGATAGAAGAATGGAATAAAAAATATCCACA 1706
           :::         ||||||::::::   ::::::||||||   ||
 578 ......AspValValAsnGlyIleLysAsnSerAspArgLysTyrLysHi 592

1707 TAAACCACTAGAAGTAGAGGAA 1728
     |      |||   :::   |||
 592 sAspAsnLeuThrThrThrGlu 599
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-912-4

seq_documentation_block:
; Sequence 4, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-4U1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1861 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-790-912-4

alignment_scores:
Quality: 109.00 Length: 537
Ratio: 0.441 Gaps: 27
Percent Similarity: 45.996 Percent Identity: 20.112

alignment_block:
US-09-579-383-2 x US-08-790-912-4 ..

Align seg 1/1 to: US-08-790-912-4 from: 1 to: 1861

```
  76 AAAAATAATATAAATAAT......TCATTGGGAATAATA.......... 108
     ::::::|||||||||||||         :::|||   :::
 874 GluGlyAsnIleAsnAsnValLysArgPheGlySerValAlaGlyValAl 890

 109 ...............CGGGAAAATAAAAATAAAACTCATCAAACGGAAA 142
                     |||:::::::::::::::   |||      ::::
 890 aGlyTyrLeuTrpAspArgAspSerSerGluGluArgHisAlaGlyArgL 907

 143 TACATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAA 192
     ::|||:::   :::|||   ::::::      |||:::|||   ::::::
 907 euHisAsnValLeuSerAspIleAsnValMetAsnGlyAsnAlaIleSer 923
```

```
 193 TATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAA 242
924 GlyTyrHisTyrArgGly.....MetArgIleThrAspSerTy 936

 243 TAAAAATATAAAT.....AAAAATGATAGAA 268
936 rSerAsnLysAspAsnArgValTyrLysValThrLeuGluLysAspGluV 953

 269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATT 318
953 alValThrLysGluSerLeuGluGlu.....ArgGlyThr 964

 319 ATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACA 368
965 IleLeuAspVal.....SerGlnIleAlaSerLysLysSe 976

 369 TATGATTGAT.....TCAAACCCAATGGTGTCAATTTTATATATTGCA. 411
976 rGluIleAsnSerLeuSerAlaProLysValGluThrLeuLeuThrSerT 993

 412 .....TTTGCTCGCATTAATATGTTATATGATGTATCT 444
993 hrAsnLysGluSerAspPheSerLysValLysAsp...TyrGlnAlaSer 1008

 445 AGA.....CCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGG 488
1009 ArgAlaLeuAlaTyrLysAsnIleGluLysLeuLeuProPheTyrAsnLy 1025

 489 CTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCA 538
1025 sAlaThrIleValLysTyrGlyAsnLeuValLysGluAspSerThrLeu. 1041

 539 GAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACC 588
1042 .....TyrGluLysGluIleLeuSerAlaValMetMetLysAspAsnGlu 1056

 589 TATATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAA 638
1057 ValIleThrAspIleAlaSerHisLysGluAlaAlaAsnLysLeuLeuIl 1073

 639 GCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCAC 688
1073 eHisTyrLysAspHisSerSerGluLysLeuAspLeuThrTyrGlnSerA 1090

 689 AT.....GGGAAGTTTTAC 702
1090 spPheSerLysLeuAlaGluTyrArgValGlyAspThrGlyLeuIleTyr 1106

 703 AACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTT 752
1107 ThrProAsnGln.....PheLeuGlnAsnHisSerSerIleValAsn.. 1120

 753 GTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATT...TCTGGTT 799
1121 .....GluValLeuProAspLeuLysAlaValAspTyrGlnSerGluA 1135

 800 CATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCA..... 837
1135 laIleArgAsnThrLeuGlyIleSerSerGlyValSerLeuThrGluLeu 1151

 838 .....TCTTTCTGTAAAGATGAAGAATCTCCATATAACACT.. 873
1152 TyrLeuGluGluGlnPheAlaLysThrLysGluAsnLeuAlaAsnThrLe 1168

 874 ....AAATTTTTGTCT..... 885
1168 uGluLysLeuLeuSerAlaAspAlaValIleAlaSerGluAsnGlnThrI 1185

 886 .....GAACAAATAGAAACAAATAAAGAATTACATAGG 918
1185 leAsnGlyTyrValValAspLysIleLysArgAsnLysGlu..... 1198

 919 GCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATACAGC 968
1199 .....AlaLeuLeuLeuGlyLeuThrTyrLeuGlu..... 1208

 969 AAAGGAGAAAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACTA 1018
1209 .....ArgTrpTyrAsnPheAsnTyrG 1216

1019 CAAATCCAGATATAATGGTAGATATGTACTTATCCCATTTATATTTTGGT 1068
1216 lyAspValAsnValLys...AspLeuValMetTyrHisMetAspPhePhe 1231

1069 TTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAACATAACAG 1118
1232 GlyLysGlyAsnValSer...ProLeuAspThrIleIleGluLeuGlyLy 1247

1119 AGGTGGATTTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAA 1168
1247 sSerGlyPheAsn.....AsnLeuLeuAlaLysAsnAsnValA 1260

1169 CAATACATGATAAAAATCAAAATAATAATAGGGCAGATGGTATAGGGATA 1218
1260 spAlaTyrAsnIleSerLeuAlaAsnAsnAsnAla..... 1271

1219 TGGCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGA 1268
1272 .....ThrLysAspLeuPheSerThrLeuAlaAsnTyrArgGl 1284

1269 TATTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAAGTACAAACTC 1318
1284 uValPheLeuProAsn.....LysThrAsnAsnGlnTrpP 1296

1319 CAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAGCACAATAGAT 1368
1296 heLysGluGlnThrLysAlaTyrIleValGluGluLysSerAlaIleAsp 1312

1369 GAA.....TATGTTCCAGGACT 1385
1313 GluValArgValLysGlnGluGlnAlaGlySerLysTyr..... 1325

1386 CGTTATTCCAACCATAGGGATATATTACAAACACAATGATGCTATATGGA 1435
1326 .....SerIleGlyValTyrAspArgIleThrSerAspThrTrpL 1339

1436 AAACTAGATCT 1446
1339 ysTyrArgAsn 1342
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-569-853-1

```
seq_documentation_block:
; Sequence 1, Application US/08569853
; Patent No. 5869036
;  GENERAL INFORMATION:
;    APPLICANT:  Belshe, Robert B.
;    TITLE OF INVENTION:  LIVE ATTENUATED VACCINES BASED ON cp45
;    TITLE OF INVENTION:  HPIV-3 STRAIN AND METHOD TO ENSURE ATTENUATION IN SUCH
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Paul A. Stone
;      STREET:  One Metropolitan Square- 16th Floor
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/569,853
```

```
;      FILING DATE:  08-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stone, Paul A.
;      REGISTRATION NUMBER:  38,628
;      REFERENCE/DOCKET NUMBER:  SLU 4471
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-231-5400
;      TELEFAX:  314-231-4342
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2233 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Human parainfluenza virus 3
;      STRAIN:  JS
US-08-569-853-1


alignment_scores:
            Quality:  109.00              Length:    673
              Ratio:  0.355                Gaps:     32
 Percent Similarity:  45.617   Percent Identity: 18.425

alignment_block:
 US-09-579-383-2 x US-08-569-853-1   ..

 Align seg 1/1  to: US-08-569-853-1  from: 1  to: 2233

       4 AATTTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTCTGC 53
         ||||||   ::::::   :::   |||:::::::   :::      |||::
    1270 AsnPheThrLeuAspSerLeuLysIleLeuThrProValAlaThrSerTh 1286

      54 AAAT...TCCAGAACCTTGAAAGGA.......................... 75
         :|||   |||:::   ||||||
    1286 rAsnLeuSerHisArgLeuLysAspThrAlaThrGlnMetLysPheSerS 1303

      76 ..................................AAAAATAATATAAAT 90
                                           :::|||   |||
    1303 erThrSerLeuIleArgValSerArgPheIleThrMetSerAsnAspAsn 1319

      91 AATTCATTGGGAATAATACGGGAAAATAAA......AATAAAACTCATCA 134
            |||:::         :::|||:::|||      |||      :::||
    1320 MetSerIleLysGluAlaAsnGluThrLysAspThrAsnLeuIleTyrGl 1336

     135 A..................ACGGAAATACATGAGTCTTTTTCACATCTTA 166
         |                  ::::::   |||   :::   :::||||
    1336 nGlnIleMetLeuThrGlyLeuSerValPheGluTyrLeuPheArgLeuL 1353

     167 AA...TCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT 213
         ||   :::::::::   |||   :::            :::      |||
    1353 ysGluThrThrGlyHisAsnProIleValMetHisLeuHisIleGluAsp 1369

     214 GGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGA 263
            |||      :::   :::   |||::::::::::||||||
    1370 GluCysCysIleLysGluSerPheAsnAspGluHisIleAsn........ 1383

     264 TAGAAAATCACCAAGACAAATTTTAGAG......GAGTATAAAAAAAGGA 307
                 |||::::::   ||||||      :::|||   :::   :
    1384 ..........ProGluSerThrLeuGluLeuIleArgTyrProGluSerA 1397

     308 AACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGAT 357
         :::::   |||         |||      :::      ::::::   |||
    1397 snGluPheIle.........TyrAspLysAspProLeuLysAspValAsp 1410

     358 AGAGCAAAACATATG.......................ATTGATTCAAA 383
            :::|||   |||                      ||||||   ||
    1411 LeuSerLysLeuMetValIleLysAspHisSerTyrThrIleAspMetAs 1427

     384 CCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTAT 433

         |    |||   :::      |||:::|||   :::::::::
    1427 n.TyrTrpAspAspThrAspIleIleHisAlaIleSerIleCysThrAla 1443

     434 ATGATGTATCTAGACCATTTAATGGAA...GACAAAGATTCCTATTAAGA 480
         :::      |||   :::   :::   |||:::|||
    1444 IleThrIleAlaAspThrMetSerGlnLeuAspArgAsp........... 1456

     481 AAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATTAG 530
                    |||:::|||   ::::::                    |
    1457 .............AsnLeuLysGluIleIleValIleAlaAsnAspAspA 1469

     531 ACGTATCAGAAAAGTACGTCCAGAT..............GTAATTATT 564
         ||:::::::::      ::::::                :::::::::
    1469 spIleAsnSerLeuIleThrGluPheLeuThrLeuAspIleLeuValPhe 1485

     565 CTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGAAAT 614
         |||   ::::::||||||
    1486 LeuLysThrPheGlyGly................................ 1491

     615 TGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTT.......... 654
                                  ||||||||||:::|||
    1492 ......................LeuLeuValAsnGlnPheAlaTyrThrL 1501

     655 ...........GATTTAGATGGTGTAGATATTGACTGGGAACCACATGGG 693
                      :::::::|||   |||:::   |||:::
    1501 euTyrSerLeuLysIleGluGlyArgAspLeuIleTrpAsp......... 1514

     694 AAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATT 743
            |||   ::::::   |||:::   :::::::::      :::::
    1515 ......TyrIleMetArgThrLeuArgAspThrSerHis......SerIl 1527

     744 AATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCA.... 789
         ::::::::::|||   :::::::::   :::   |||:::::::::
    1527 eLeuLysValLeuSerAsnAlaLeuSerHisProLysValPheLysArgP 1544

     790 ........................ATTTCTGGTTCATCAAATGCTGCA 813
                                     |||   |||   :::::::|||:::
    1544 heTrpAspCysGlyValLeuAsnProIleTyrGlyProAsnThrAlaSer 1560

     814 TTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTCC 863
         :::   ::::::   :::   |||:::|||:::   :::   :::
    1561 GlnAspGlnIleLysLeuAlaLeuSerIleCysGluTyrSerLeuAspLe 1577

     864 ATATAACACTAAATTTTTGTCT............................ 885
         :::      ::::::|||:::
    1577 uPheMetArgGluTrpLeuAsnGlyValSerLeuGluIleTyrIleCysA 1594

     886 ............................GAACAAATAGAAACAAATAAA 906
                                       :::|||         ::::::
    1594 spSerAspMetGluValAlaAsnAspArgLysGlnAlaPheIleSerArg 1610

     907 GAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTT...ATTAA 953
         :::|||      ::::::::::::   :::   :::::::|||      ||
    1611 HisLeuSerPheValCysCysLeuAlaGluIleAlaSerPheGlyProAs 1627

     954 TATTTTTAAT...ACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
         |:::::::|||   |||      |||::::::||||||      ::::::
    1627 nLeuLeuAsnLeuThrTyrLeuGluArgLeuAspLeu......LeuLysG 1642

    1001 CATAC.........AATTTAGAAACTACAAATCCAGATATAATGGTAGAT 1041
           |||         |||:::::::      :::|||   :::
    1642 lnTyrLeuGluLeuAsnIleLys...GluAspProThrLeuLysTyrVal 1657

    1042 ATGTACTTATCCCATTTATATTTTGGTTTAAAATATAACATCACAATCAT 1091
         :::      :::   |||      ::::::      :::::::|||   ::
    1658 GlnIleSerGlyLeuLeuIleLysSerPheProSerThrValThrTyrVa 1674

    1092 ATTAGGTTTTTCATTAGAACATAACAGA...GGTGGATTTAGTCCC.... 1134
         :         ::::::::::::   |||      |||:::||||||
    1674 lArgLysThrAlaIleLysTyrLeuArgIleArgGlyIleSerProProG 1691

    1135 ........................GAAAATAAAGAATTATTAGAATTG 1158
                                   |||:::::::::::::|||:::
```

```
1691 luValIleAspAspTrpAspProValGluAspGluAsnMetLeuAspAsn 1707

1159 GTAGGAAAAACAATACATGATAAAAATCAAAATAATAATAGGGCAGATGG 1208
     :::   |||||||||:::|||:::   ::::::::::|||::::::::::
1708 IleValLysThrIleAsnAspAsnCysAsnLysAspAsnLysGlyAsnLy 1724

1209 TATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAACTGGATCAT 1258
      |||
1724 sIle..................................................... 1725

1259 TCGATGTAGATATTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAA 1308
                       :::|||:::|||               :::
1726 ...................AsnAsnPheTrpGlyLeuAlaLeuLysAsn 1735

1309 GTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAG 1358
        |||:::   |||      :::||||||::::::   :::|||   ::
1736 TyrGlnValLeuLysIleArgSerIleThrSerAspSerAspAspAsnAs 1752

1359 CACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATAT 1408
     :   :::|||      :::   ||||||::::::|||      |||   |
1752 pArgLeuAspAlaAsnThrSerGlyLeuThrLeuProGlnGlyGlyAsnT 1769

1409 ATTACAAACACAATGATGCTATATGGAAAACTAGATCTTATTCA...... 1452
     ||   :::|||:::      ::::::      :::|||   |||
1769 yrLeuSerHisGlnLeuArgLeuPheGlyIleAsnSerThrSerCysLeu 1785

1453 ...................ATTCATGCACCTGGTGTAGACAGATATGA 1481
                         |||            |||::::::   ::
1786 LysAlaLeuGluLeuSerGlnIleLeuMetLysGluValAsnLysAspLy 1802

1482 ATGGGACTTG.....................GTCAAAGTATGCTATG 1507
     :      |||                     :::   :::||||||:
1802 sAspArgLeuPheLeuGlyGluGlyAlaGlyAlaMetLeuAlaCysTyrA 1819

1508 AAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGACTATAAA 1557
     ::          |||   |||::::::||||||||||:::      :::
1819 spAlaThrLeu...GlyProAlaValAsnTyrTyrAsnSerGlyLeuAsn 1834

1558 GAAAGCTCTATTATT.................ATATGGAAAGGGGAACC 1589
        ::::::::::|||                |||:::   :::|||
1835 IleThrAspValIleGlyGlnArgGluLeuLysIleProSerGluVa 1851

1590 ATATTTAATTAAATGGTGGCAACAAGGACCTCCGGAAGGTCAGGCACTAG 1639
        |||:::                        |||:::   |||
1851 lSerLeuVal.........................GlyLysLysLeuG 1859

1640 AGTCATACACAAAACTA 1656
       :::   |||::::::
1859 lyAsnValThrGlnIle 1864
```

OM of: US-09-579-383-2 to: PIR_68:* out_format : pfs

Date: Mar 21, 2002 4:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09579383/runat_20032002_151239_29320/app_query.fasta_1.1849
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09579383_@CGN1_1_89 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-579-383-2
Query length: 1764
Database: PIR_68:*
Database sequences: 219241
Database length: 76174552
Search time (sec): 99.360000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
pir1:S34960	-	158.50	184.71	0.0045	590	! NADH dehydrogenase (ubiquinone)
pir2:T18482	+	150.50	163.60	0.0165	2437	! hypothetical protein C0705c -
pir2:S61166	+	149.00	174.61	0.0191	511	! probable membrane protein YDR37
pir2:S72284	-	148.50	168.86	0.0213	960	! DNA-directed RNA polymerase (EC
pir2:T43916	+	148.00	166.33	0.0233	1215	! chitinase A [imported] - Pyroc
pir2:S60651	+	145.50	169.67	0.0328	563	! chitinase precursor - Serratia
pir2:A38221	+	145.00	169.98	0.0352	504	! chitinase (EC 3.2.1.14) MF1 - m
pir2:H71611	+	144.00	158.29	0.0437	1817	! probable secreted protein PFB0
pir2:T24076	+	143.00	168.64	0.0474	444	! hypothetical protein R09D1.5 -
pir1:I30010	-	142.50	167.04	0.0514	502	! NADH dehydrogenase (ubiquinone)
pir2:A25090	+	142.50	166.13	0.0517	561	! chitinase (EC 3.2.1.14) - Serra
pir2:B71621	+	139.50	151.04	0.0876	2295	! probable membrane associated p
pir2:T30933	+	138.50	156.23	0.0981	1054	! chitinase (EC 3.2.1.14) A - Ps
pir2:B71612	+	138.00	144.77	0.1132	3973	! hypothetical protein PFB0555c
pir1:QQUTC5	-	136.50	158.60	0.1292	590	! NADH dehydrogenase (ubiquinone)
pir2:T03884	+	136.50	148.51	0.1375	2025	! hypothetical protein F07G11.9
pir2:S07915	+	136.00	152.57	0.1442	1146	! RF2 protein - yeast (Kluyverom
pir2:T41863	+	135.50	157.96	0.1500	552	! chitinase chi-A orf126 - Bombyx
pir2:G71609	+	135.50	145.59	0.1618	2500	! hypothetical protein PFB0650w
pir2:B26696	-	135.00	159.17	0.1601	443	! hypothetical protein 1 (CYb-COI
pir2:T18460	+	133.50	152.19	0.2076	836	! hypothetical protein C0510w - m
pir2:T18497	+	133.50	142.72	0.2200	2657	! hypothetical protein C0780w -
pir2:T18501	+	133.50	140.72	0.2227	3394	! hypothetical protein C0760c -
pir2:T18414	+	133.00	140.82	0.2393	3119	! protein g377 - malaria parasit
pir2:S58360	+	132.50	149.81	0.2435	967	! lantibiotic Pep5 biosynthesis p
pir2:A45624	+	131.50	152.97	0.2761	569	! trophozoite cysteine proteinase
pir2:A45597	+	131.00	140.80	0.3198	2339	! DNA-directed RNA polymerase (E
pir2:T29275	+	130.50	143.93	0.3373	1484	! hypothetical protein T01C4.1 -
pir2:T28160	+	130.00	139.04	0.3737	2510	! hypothetical protein - malaria
pir2:T18440	+	130.00	134.17	0.3850	4550	! hypothetical protein C0425w -
pir1:A44267	+	129.50	151.32	0.3727	521	! protein-tyrosine-phosphatase (E
pir2:G72865	+	129.50	150.86	0.3737	551	! chitinase - Autographa californ
pir2:T24085	+	129.00	152.26	0.3984	432	! hypothetical protein R09D1.11 -
pir2:T18461	+	129.00	147.54	0.4100	768	! hypothetical protein C0505c - m
pir2:T18429	+	129.00	140.99	0.4268	1711	! hypothetical protein C0345w -
pir2:G70163	+	129.00	139.06	0.4319	2166	! hypothetical protein BB0512 -
pir2:T00323	+	128.50	146.31	0.4442	831	! chitinase (EC 3.2.1.14) B - Clo
pir2:S41649	+	128.50	139.73	0.4624	1855	! DNA polymerase - malaria paras
pir2:T24074	+	128.00	151.01	0.4640	435	! hypothetical protein R09D1.3 -
pir2:T09079	+	127.50	135.45	0.5487	2708	! probable chloroquine resistanc
pir2:T23643	+	127.00	149.04	0.5429	479	! hypothetical protein M01B2.6 -
pir2:E70118	+	126.50	145.76	0.5955	665	! flagellar hook-associated pr
pir2:C71622	+	126.50	136.83	0.6290	1979	! hypothetical protein PFB014
pir2:T18427	+	126.50	131.65	0.6492	3724	! hypothetical protein C0335c
pir2:F82885	+	126.50	129.77	0.6567	4688	! hypothetical protein UU482

seq_name: pir1:S34960

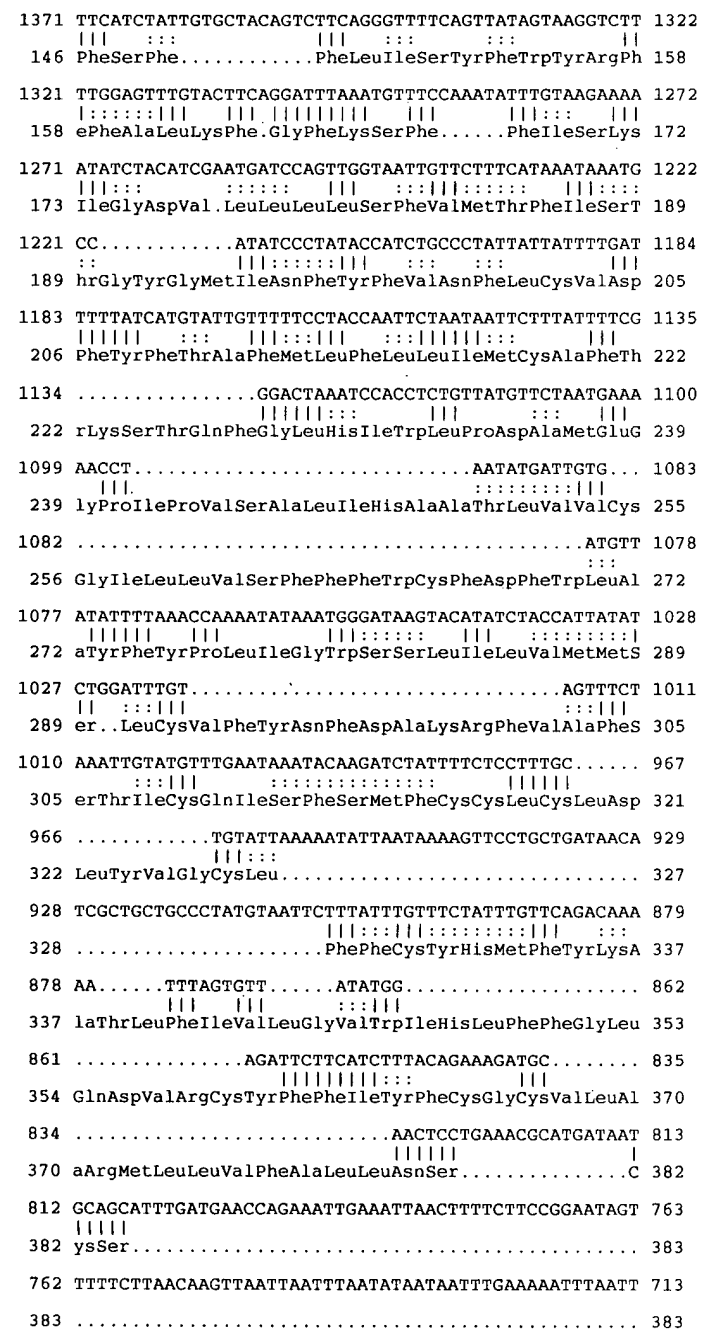
seq_documentation_block:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Crithidia oncopelti mitochondr
C;Species: mitochondrion Crithidia oncopelti
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C;Accession: S34960
R;Maslov, D.A.; Horvath, A.; Gwang II, K.; Kolesnikov, A.A.
submitted to the EMBL Data Library, October 1990
A;Reference number: S34958
A;Accession: S34960
A;Molecule type: DNA
A;Residues: 1-590 <MAS>
A;Cross-references: EMBL:X56015; NID:g12879; PIDN:CAA39492.1; PID:g12882
C;Genetics:
A;Gene: ND5
A;Genome: mitochondrion
A;Genetic code: SGC6
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylatio

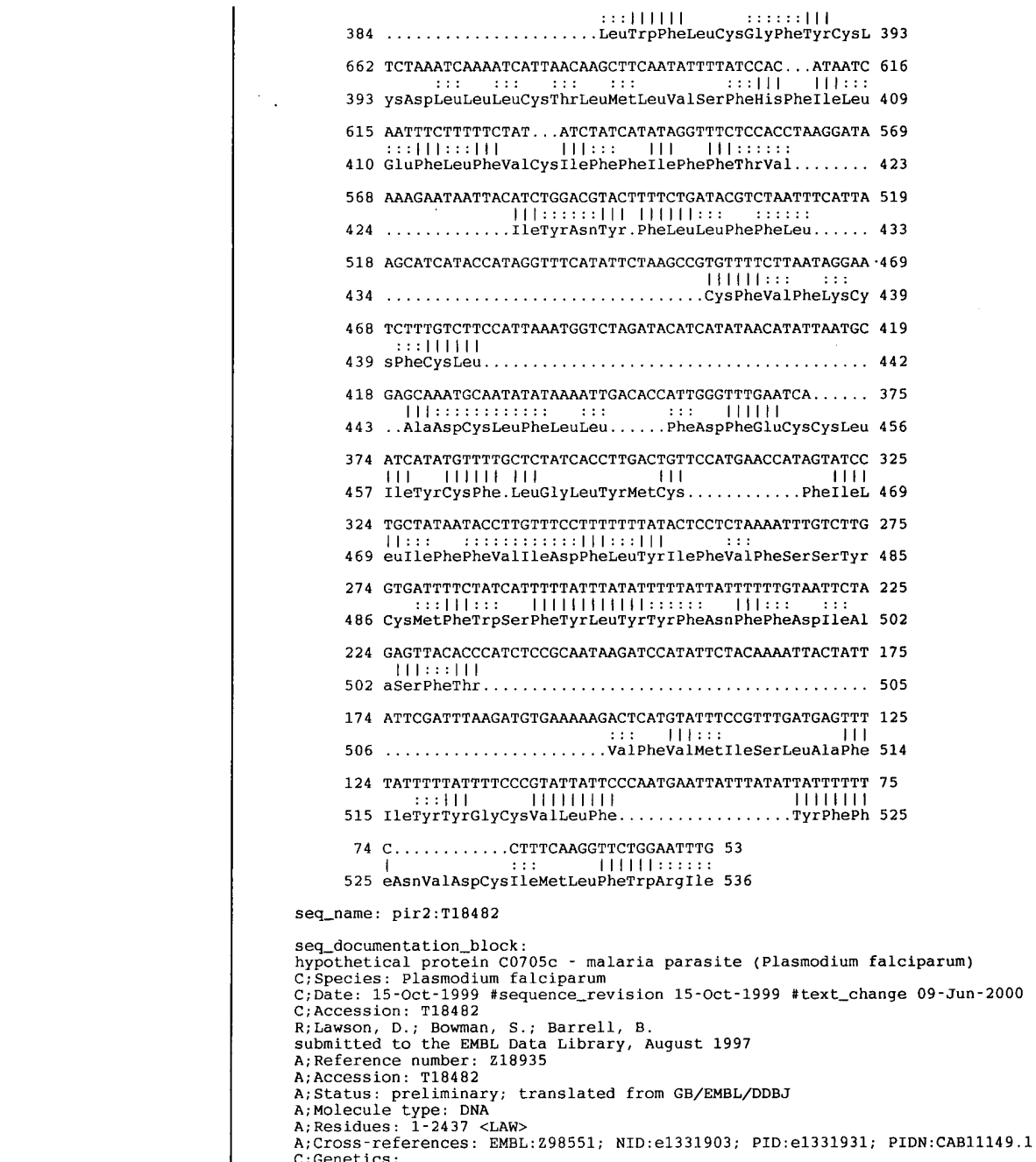
alignment_scores:
Quality: 158.50 Length: 681
Ratio: 0.576 Gaps: 41
Percent Similarity: 40.382 Percent Identity: 21.145

alignment_block:
US-09-579-383-2/rev x S34960 ..

Align seg 1/1 to: S34960 from: 1 to: 590

```
1743 TTCTTGTTCATATTGTTCCTCTACTTCTAGTGGTTTATGTGGATATTTTT 1694
     |||:::|||:::||||||:::   |||   :::|||:::      |||||
   4 PhePhePheMetLeuPhePheLeuPheGlyPhePheLeuGlyThrPhePh 20

1693 T..............ATTCCATTCTTCTATCCC..............  1675
     |               :::   |||:::
  20 eLeuGlyArgHisLeuLeuSerPheTrpLeuSerIleLeuMetThrValP 37

1674 ...................TGGACATTTGGATGCATCTAGTTTTGTGTA 1645
                        ::::::|||   ||||||      |||:::
  37 heLeuValMetValThrMetPheSerPhePheCysIleSerValCysLeu 53

1644 TGACTCTAGTGC...CTGACCTTCCGGAGGTCCTTGTTGCCACCATTTAA 1598
           |||         |||         |||:::      :::
  54 TyrGlyTyrCysTyrTyrAspPheCysLeuIleLeuMetLeuAspLeuCy 70

1597 TTAAATATGGTTCCCCTTTCCATATAATAATAGAGCTTTCTTTATAGTCA 1548
        |||::::::   |||      |||:::      :::|||||||:::
  70 sPheIlePheMetSerPheTyrCysAsnGlyPheTyrLeuPheIleLeuP 87

1547 GTGTTATA...ATAATGGGCTGCTTTCCCATCGCATAT...TTTTTCATA 1504
       :::|||   ::::::   ||||||   :::   |||   |||:::
  87 heLeuIleAspLeuValPheCysPheIleLeuPheTyrAlaPheTyrTyr 103

1503 GCATACTTTGACCAAGTC......CCATTCATATCTGTC.........TA 1469
        |||:::|||   :::         |||:::::::::         ::
 104 MetTyrTyrAspLeuMetLeuLysArgPhePheAsnIlePheTrpTrpPh 120

1468 CACCAGGTGCATGAATTGAATAAGATCTAGTTTTCCATATAGCATCATTG 1419
     ::::   |||||||||   :::      |||:::   |||
  120 eValLeuCysMetAsnPhePheIleLeuSerTyrAspTyr.......... 133

1418 TGTTTGTAATATATCCCTAT...GGTTGGAATAACGAGTCCTGGAACATA 1372
               :::      |||   ||||||      |||      ::::::
  134 ........LeuThrAlaTyrCysGlyTrp......GluLeuLeuGlyLeu 145
```





```
seq_name: pir2:T18482

seq_documentation_block:
hypothetical protein C0705c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18482
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18482
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2437 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331931; PIDN:CAB11149.1
C;Genetics:
A;Map position: 3
A;Introns: 1397/1
```

A;Note: C0705c

alignment_scores:
Quality: 150.50 Length: 655
Ratio: 0.514 Gaps: 38
Percent Similarity: 44.733 Percent Identity: 19.847

alignment_block:
US-09-579-383-2 x T18482 ..

Align seg 1/1 to: T18482 from: 1 to: 2437

```
  76 AAAAATAATATAAATAATTCATTGGGAATA...ATACGGGAAAATAAAAA 122
     ||||||:::::::::::::         :::   ||||||:::::::|||||
1556 LysAsnLysValAspLysIleTyrGluValGluIleArgLysGluLysAs 1572

 123 TAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTTAAATCGA 172
     ||||   ::::::::::::|||:::::   :::|||:::          |
1572 nLysLysAsnLysAsnArgIleAsnLys...TyrSerTyr.........A 1585

 173 ATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT......GGG 216
     |||||::::::::::::                        |||
1585 snAsnLysGluTyrIle..............LeuLysAspLeuArgArg 1596

 217 TGTAACTCTAGAATTACAAAAAATAATAAA..................... 246
     |||::::::::::|||   ||||||||||||
1597 CysGlnLysLysIleLysLysAsnAsnLysIleIleIleLysIleAsnAs 1613

 247 ..............AATATAAATAAAAATGATAGAAAATCACCAAGAC 280
                   |||:::|||:::        :::      |||:
1613 nLysCysAsnMetTyrAsnMetAsnAsnIleTyrSerArgIleLeuArgA 1630

 281 AAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATAC 330
     :::::|||::::::   ||||||||||||:::   ::::::::::::|||
1630 snPheLeuAsnLysAsnLysLysArgLysAsnLysLeuPheSerSerTyr 1646

 331 TATGGTTCATGGAACAGTCAAGGTGATAGAGCA......AAACATATGAT 374
       :::     |||:::       |||:::      ||||||:::||
1647 ValAsnIleLeuAsnAspCysLeuHisArgThrIleArgLysHisIleIl 1663

 375 TGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTA 424
     |   :::|||                  :::|||
1663 eLysAsnAsn.................PheIle................ 1668

 425 ATATGTTATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTA 474
                              :::|||::::::|||:::
1669 .........................TyrAsnAsnLysGlnAsn...... 1674

 475 TTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGA 524

1674 ..................................................... 1674

 525 AATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCT 574
                 :::|||:::      |||:::      |||   ||||
1675 .............AsnLysIleValGlnAspIleLysGluLeuTyrSerL 1687

 575 TAGGTGGAGAAACCTATATGATAGATATAGAA..................... 606
     ||::::::         :::   |||||||||
1687 euSerSerValHisAsnIleArgAspIleGluCysGlyTyrGlnGluArg 1703

 607 ......AAAGAAATTGATTATGTGGATAAAATATTG.............AA 638
           ||||||:::         |||   |||:::             ||
1704 MetArgLysGluMetLeuLysPheAspIleIleIleAspHisLysGluLy 1720

 639 GCTTGTTAATGATTTTGAT.........TTAGATGGTGTAGATATTGAC. 678
     ||||:::|||:::|||:::         :::|||:::::::::::::
1720 sLeuIleAsnSerPheAsnPheValCysValAspAsnIleSerLeuGlyL 1737

 678 ..................................................... 678

1737 euHisAsnAsnAsnAsnAsnAsnLysAlaLeuGluAsnLysGlyLysSer 1753

 679 ......TGGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTT 722
           ::::::|||:::|||:::       |||:::|||
1754 GlyAlaTyrLysProTyrGlySerAsnLysAsnIleAsn........... 1766

 723 TTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGG 772
                   |||   ||||||   ::::::|||:::|||
1767 ...............LysAsnIleAsn...IleAsnLysAsnIleAsnI 1777

 773 AAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGC 822
       :::|||   |||:::|||:::  :::   ||| ::::::   :::::
1777 leAsnLysAsnIleAsnIleAsnLysAsnIleAsn.IleAsnLysAsnVa 1793

 823 GTTTCAGGAGTTGCATCTTTCTGTAAAGATG.........AAGAATCTCC 863
     :   :::   :::::::::::::::::::         ||||||::::
1793 lAsnIleAsnLysAsnValAsnIleAsnGlnAsnIleAsnLysAsnIleA 1810

 864 ATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTAC 913
     ::|||::::::|||:::   ::::::|||   :::::::::::::|||
1810 snIleAsnIleAsnIleAsnIleAspLysGlySerAsnMetAsnAsnPro 1826

 914 ATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTT.............. 948
       ::::::|||:::      ::::::::::::
1827 CysAsnHisGlnAsnLysAsnArgGluAspIleHisTyrAsnAsnHisTy 1843

 949 ..........ATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTG 988
             |||   :::   |||      ::::::|||   :::|||:
1843 rLysHisMetIleTyrPheHisAsnLysHisAsnAspLysSerGluLeuL 1860

 989 TA.................................................. 990
     ::
1860 euLysLysAsnSerGlyGlyLeuGluTyrAsnHisGluGluArgLysTyr 1876

 990 ..................................................... 990

1877 LysGluLysIleTyrSerCysLysArgAsnPheGluIleGluAsnIleAs 1893

 991 ....TTTATTCAA.....................ACATACAATTTAGAAA 1015
         ||||||                        |||:::   :::
1893 nTyrPheIleTyrArgAsnIleTyrArgProIleLysTyrHisTyrAspI 1910

1016 CTACAAATCCAGATATAATGGTAGATATGTACTTATCCCAT......... 1056
       ||||||      |||::::::::::   :::|||   :::
1910 leThrAsn...LysIleLeuIleAsnSerPheLeuIleAsnMetAsnLeu 1925

1057 ...TTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTC 1103
       :::||||||          |||   :::      ||||||      ::
1926 LeuIleTyrPhe.........TyrLeuLeuLysCysIleLeuPheAspAs 1939

1104 ATTAGAACATAACAGAGGTGGA..............TTTAGTCCCGAAA 1138
     :::::::   |||   :::                  |||:::   :::|
1939 pMetAspAspAsnTyrAsnPheLeuCysAsnIleLeuPheAsnLysArgA 1956

1139 ATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAA 1188
     |||||            :::::::::|||   |||::::::   :::
1956 snLys............IleIleAsnLys...IleGluAsnAspAspTyr 1967

1189 AATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAAGA 1238
     |||:::|||   :::|||               |||:::|||:::::::
1968 AsnAspAsnCysCysAsp..............HisIlePheValAsnLy 1979

1239 ACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACA...AATA 1285
     :           ::::::      |||:::   ||||||:::   |||:
1979 s............AsnLysHisMetValGlnTyrPheLeuAsnLysAsnP 1992

1286 TTTGGAAACATTTAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATA 1335
     :::::         |||                  |||:::      :::
1992 heTyrThrSerCysAsn..............IleLysAsnLysCysVal 2003

1336 ACTGAAAACCCT.........GAAGACTGTAGCACAATAGATGAATATGT 1376
       :::|||               |||   |||||||||
2004 AspLysAsnLysLeuTyrTyrThrAspIleSerThrIle........... 2016

1377 TCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAACACAATGATG 1426
```

```
                                      ::::::|||   |||
    2017 ..........................PhePheTyrIleHis..... 2021

    1427 CTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGA 1476
                               ::::::|||::::::|||
    2022 ...................PheLysIleAsnValPro........... 2027

    1477 TATGAATGGGACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAA 1526
                  |||:::::::::   :::          |||   ||
    2028 .........LysLeuLeuArgIlePhePhe...........AspHisLy 2037

    1527 AGCAGCCCATTATTATAACACTGACTATAAAGAAAGCTCTATTATT.... 1572
         |:::::::::|||||||||      |||:::   :::|||::::::
    2037 sValValAsnTyrTyrAsnAspIleTyrArgLeuThrSerLeuLeuTyrP 2054

    1573 ..........................................ATATGG 1578
                                                    :::
    2054 heSerLeuLeuAsnLeuAsnTyrIlePheLeuIlePheSerThrLeuCys 2070

    1579 AAAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCTCCG..... 1623
         :::         :::      |||   |||:::   |||   |||
    2071 ArgProThrIlePheGlnCysLysLysTrpLysGlyGlyAsnProSerTh 2087

    1624 ....GAAGGTCAG 1632
             |||||||||
    2087 rAlaGluGlyGln 2091

seq_name: pir2:S61166

seq_documentation_block:
probable membrane protein YDR371w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein D9481.7
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 17-Mar-2000
C;Accession: S61166
R;Ding, H.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of S. cerevisiae cosmid 9481.
A;Reference number: S61159
A;Accession: S61166
A;Molecule type: DNA
A;Residues: 1-511 <DIN>
A;Cross-references: EMBL:U28373; NID:g849184; PID:g849192; GSPDB:GN00004; MIPS:YDR371w
A;Experimental source: strain S288C (AB972)
C;Genetics:
A;Gene: MIPS:YDR371w
A;Map position: 4R
C;Superfamily: Serratia marcescens chitinase
C;Keywords: transmembrane protein
F;18-34/Domain: transmembrane #status predicted <TMM>


alignment_scores:
           Quality:  149.00            Length:    599
             Ratio:  0.598              Gaps:     28
 Percent Similarity:  41.569   Percent Identity: 18.698

alignment_block:
 US-09-579-383-2 x S61166   ..

 Align seg 1/1  to: S61166  from: 1  to: 511

     140 AAATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTA 189
         ::::::   |||         ::::::::::::::::::::::|||
       7 GlnHisArgSerLysSerSerLeuValSerHisLeuLeuIleLeuLeuIl 23

     190 GAATATGGATCTTATTGCGGAGATGGGTGT...AACTCTAGAATTACAAA 236
           :::   :::|||   ||||||  |||      ::::::|||   ||
      23 ePheIleThrIleIleIleGluMet..CysLeuTyrAsnLysIlePheLy 39

     237 AAATAATAAA.................AATATAAATAAAAATGATAGAA 268
         ||||::::::                 |||:::|||::::::   ::::
      39 sAsnGlnArgSerAspAspIleArgAspAsnPheAsnAsnGlyGlyHisA 56

     269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAA...AGGAAACAAGGT 315
         ::   |||   ::::::   ::::::            |||   ::::::
      56 rgValProSerAsnValGlnAsnHisGlyThrHisIleArgAspGluAla 72

     316 ATTATAGCAGGA...TACTATGGTTCATGGAACAGTCAAGGTGATAGAGC 362
         :::|||:::|||   ||||||::::::|||:::             |||
      73 PheIleSerGlyValTyrTyrSerAsnTrpSerProTyrLysProArgPh 89

     363 AAAACATATGATTGATTCAAAC...CCAATGGTGTCAATTTTATATATTG 409
                       |||   |||      :::|||||||   :::|||   |
      89 eHisPheProHisAspIleAsnLeuLysGlnValSerHisIleTyrTyrA 106

     410 CATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGA 459
         |||||   :::|||||||
     106 laPhePheLysIleAsnSer..............................  112

     460 AGACAAAGATTCCTATTAAGAAAACACGGCTTAGAA.............. 495
                         |||      |||:::|||
     113 ..................ArgThrGlyGlyIleGluAsnThrAspSerTr 123

     496 .....................TATGAAACCTATGGTATGATGCTTAATG 523
                              |||:::::::   ::::::      :::|
     123 pSerAspLeuGluMetAsnLeuTyrLysSerLeuAlaIleLysAsnSerG 140

     524 AAATT.....................AGACGTATCAGAAAAGTACGTCCA... 552
         ||:::                    :::   ::::::::::::   |||
     140 luLeuIleLysGluSerSerAsnAsnSerValGlnAsnIleLeuProLeu 156

     553 ..............................................GA 554

     157 GlyCysIleGlyGluLeuPheTyrLeuLysAsnThrCysSerAspLysLy 173

     555 TGTAATTATTCTTTTATCCTTAGGTGGA............GAAACCTATA 592
               ::::::::::|||:::||||||            |||::::::
     173 sPheLysValIleMetSerIleGlyGlyTrpSerAspSerGluAsnPheL 190

     593 TGATAGATATAGAAAAAGAA.........ATTGATTATGTGGATAAAATA 633
           |||   |||:::   :::            ::::::||||||:::
     190 ysIleIleIleLysAspAspLysLeuLeuGlnAsnPheValAspSerSer 206

     634 TTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGGA 683
         ::::::   :::       :::   :::||||||:::|||:::||||||||
     207 ValGluThrMetPheArgLeuGlyPheAspGlyIleAspLeuAspTrpGl 223

     684 ACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTATT 733
         |       |||              |||:::   :::      :::   |
     223 uPheProGly..............AsnAsnGluSerGluProArgGlyT 235

     734 ATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTA 783
         ||:::||||||:::::::::||||||                  ||||||
     235 yrLeuLysLeuValArgMetLeuArgLeu..............LysLeu 246

     784 ATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGT 833
           |||:::
     247 AsnSerLeu.................................... 249

     834 TGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTGT 883
                                    ::::::::::::::::
     250 ..............................GluSerGlnIlePheG 255

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         ::::::::   |||   :::   :::|||      ||||||      :::
     255 lyLysArgThrGluAspHisPheGlnLeuSerIleAlaAlaProAlaPhe 271

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
         :::         |||         ::::::         :::   :::||
     272 LysAspLysLeuPheTyrLeuProIleThrGluIleAspGlnTyrValAs 288

     984 TCTTGTATTTATTCAAACATACAATTTA..............GAAACTA 1018
         |       ::::::|||||:::                    |||||||
     288 pTyrTrpAsnMetMetThrTyrAspTyrTyrGlySerTrpSerGluThrT 305
```

```
1019 CAAATCCAGATATAATGGTAGATATGTACTTATCCCATTTATAT..... 1062
     ||:::                  |||   |||:::|||:::
 305 hrGly.....................TyrHisSerAsnLeuPheSerGlu 314

1063 .................TTTGGTTTAAAATATAACATC........... 1083
                      |||::::::   |||   |||
 315 ThrGluLeuAsnGlyAsnPheAlaMetHisTyrMetIleAspArgPheGl 331

1084 .............ACAATCATATTAGGT.....................T 1099
                    ::::::||||||                     |
 331 yValAsnSerArgLysLeuValLeuGlyMetAlaAlaTyrGlyArgSerP 348

1100 TTTCATTAGAACATAACAGAGGTGGATTTAGTCCC......GAAAATAAA 1143
     ||   ::::::   |||:::      |||:::|||      :::|||
 348 heHisIleLysAspAsnLys......PheGluProPheAsnGlnAsnThr 362

1144 GAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAA 1193
       |||::::::   :::   |||   :::        :::::::::
 363 ValLeuIleAsnLysIlePheLysGlyValGlyLysProThrLysGluIl 379

1194 TAATAGGGCAGATGGT...ATAGGGATATGGCATTTATTTATGAAAGAAC 1240
      ::::::|||||||||      |||||||||
 379 eAspLysAlaAspGlyLysGluGlyIleTrp..................... 389

1241 AATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGG 1290

 389 ..................................................... 389

1291 AAACATTTAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGA 1340
                |||      :::   ||||||
 390 ............ProTyrLysAsnLeuProLys.................. 396

1341 AAACCCTGAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTA 1390
                    :::||||||:::::|||   |||      |||
 397 .............IleGlyThrIleGluGlnTyrAspProLysTyrValS 409

1391 TTCCA.....................ACCATAGGGATATATTACAAACAC 1419
                               :::|||   |||   |||   :::
 409 erAlaTyrCysPheAspGluLysAsnSerIlePheIleSerTyrAspAsn 425

1420 AATGATGCTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTGT 1469
     :::   ::::::           :::|||      |||
 426 ThrLysSerValLysThrLysAlaGluTyrValThrHisAsnAsnLeuGl 442

1470 AGACAGATATGAATGGGACTTGGTCAAAGTATGCTATGAAAAA....... 1512
           :::   |||:::        :::|||   :::
 442 yGlyGlyPheTrpTrpGluSerCysGlyGluAlaTyrAlaAsnGluSerA 459

1512 ..................................................... 1512

 459 rgSerLeuIleAsnAlaPheAsnGluGlyLeuHisPheAsnValSerSer 475

1513 ............ATATGCGATGGGAAAGCAGCCCATTATTATAACACTGA 1550
                 :::   |||   ::::::      ||||||   :::
 476 LysProSerIlePheGlnAspValArgValLysLysTyrTyrLeuAsnLy 492

1551 CTATAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTA 1596
       |||   ::::::::::::::           :::|||||||||
 492 sTyrGlyAspGlyGlyPheLeu............SerProTyrLeu 503
```

seq_name: pir2:S72284

seq_documentation_block:
DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain - Plasmodium falciparum plastid
N;Alternate names: RNA polymerase rpoC2
C;Species: plastid Plasmodium falciparum
C;Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 20-Jun-2000
C;Accession: S72284
R;Wilson, R.J.M.; Denny, P.W.; Preiser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; Whyt
J. Mol. Biol. 261, 155-172, 1996
A;Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium fa
A;Reference number: S72277; MUID:96346169
A;Accession: S72284
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-960 <WIL>
A;Cross-references: EMBL:X95275; NID:g1171583; PIDN:CAA64574.1; PID:g1171589
A;Note: biosynthesis of this protein involves a -1 frameshift in the codon for residu
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C;Genetics:
A;Gene: rpoC2
A;Genome: plastid
A;Note: this apparently degenerate plastid is referred to as the apicoplast
C;Superfamily: DNA-directed RNA polymerase beta chain
C;Keywords: nucleotidyltransferase; plastid; transcription

alignment_scores:
Quality: 148.50 Length: 661
Ratio: 0.488 Gaps: 43
Percent Similarity: 45.991 Percent Identity: 22.995

alignment_block:
US-09-579-383-2/rev x S72284 ..

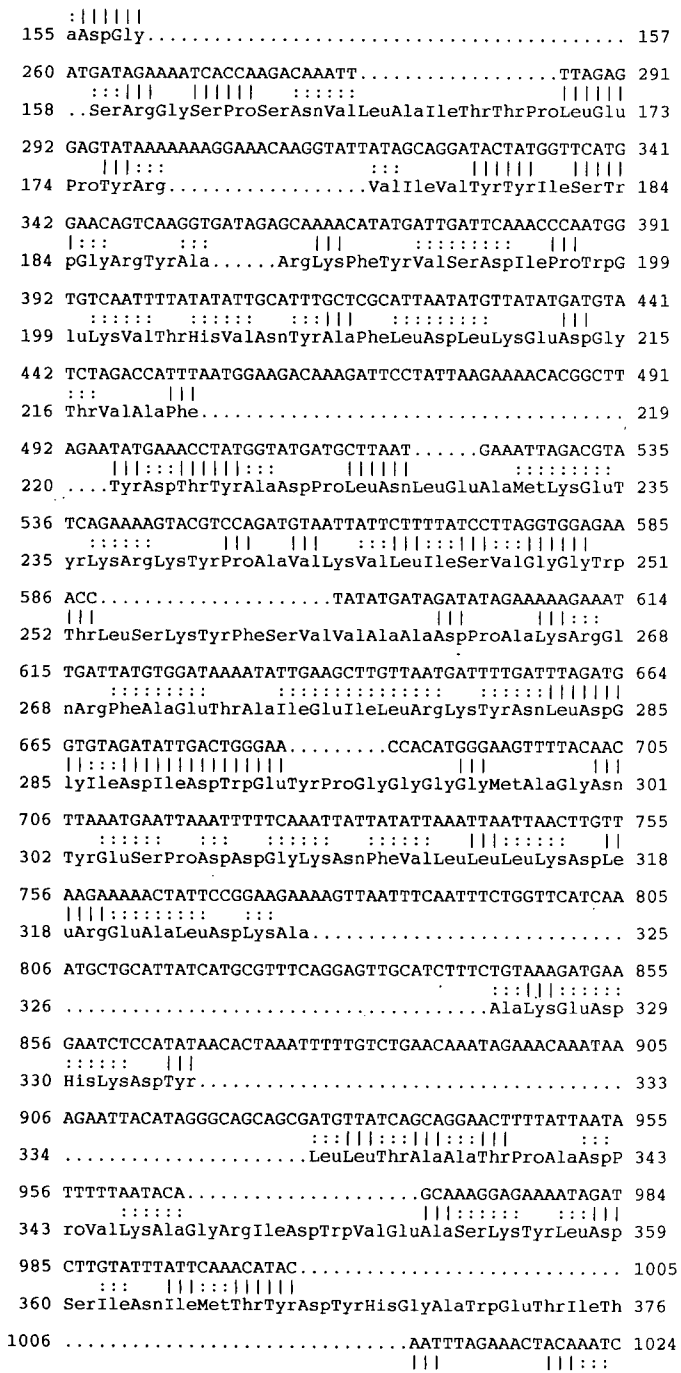
Align seg 1/1 to: S72284 from: 1 to: 960

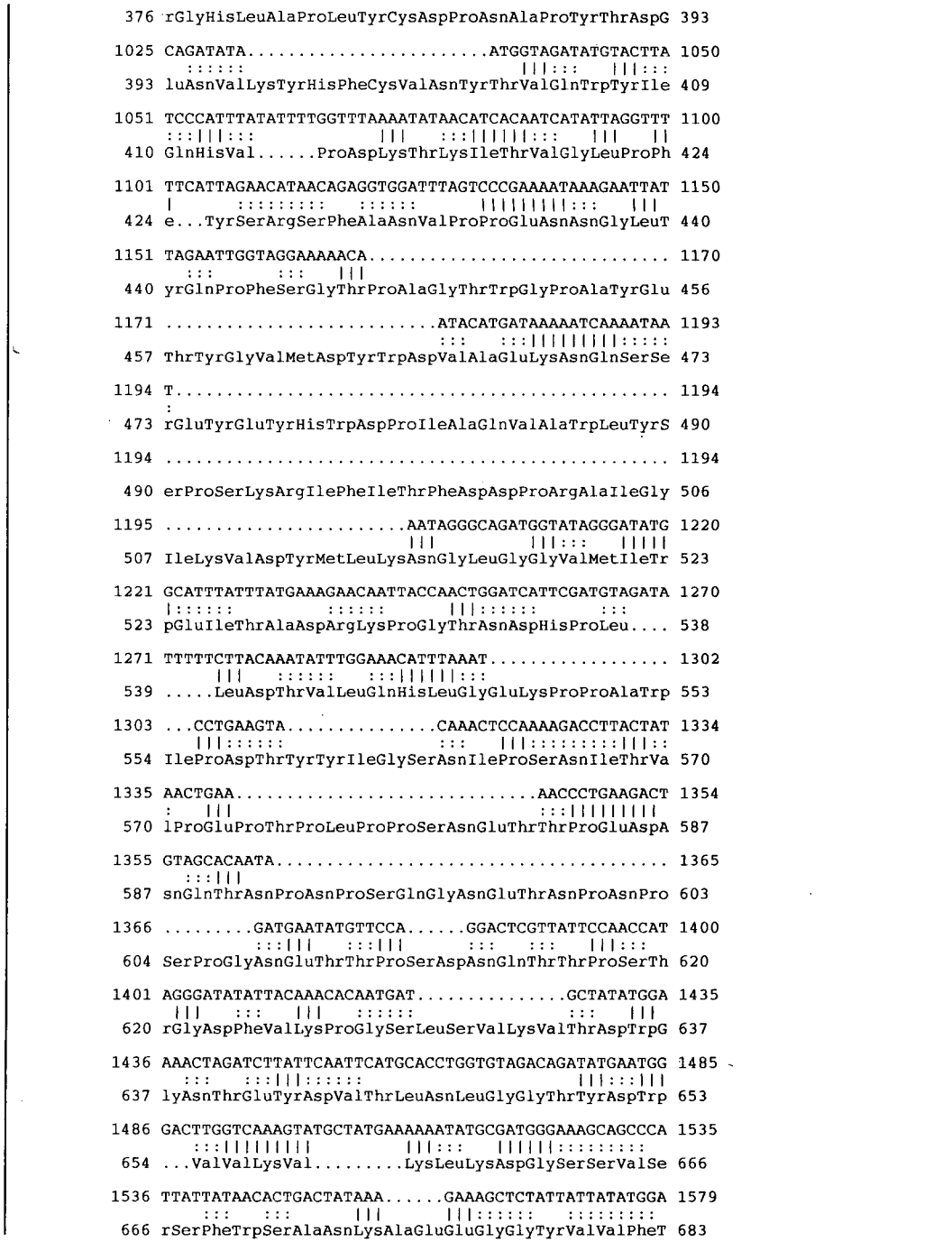
```
1561 TTTCTTTATAGTCAGTGTTATAATAATGGGCTGCTTTCCCATCGC..... 1517
     |||   :::   :::   ||||||      :::|||:::   :::
   3 PheTyrPhePheAsnLysTyrAsnLeuLysIleLeuGluLysLysLeuLe 19

1516 .ATATTTTTTCATAGCATACTTTGACCAAGTCCCATTCATATCTGTCTAC 1468
      |||:::|||                          :::|||   :::
  19 uIleIlePhe.....................LysTyrAsnIleSerPheL 29

1467 ACCAGGTGCATGAATTGAATAAGATCTAGTTTTCCATATAGCATCATTGT 1418
          :::||||||||| :::      ::::::   |||||||:::
  29 ysIleLeuHisGluLeu.LeuTyrLeuGlyTyrGluTyrSerPhe..... 43

1417 GTTTGTAATATATCCCTATGGTTGGAATAACGAGTCCTGGAACATATTCA 1368
           :::|||   |||:::   |||   ::::::::::::   ||||
  44 .......LeuTyrAsnTyrSerLeuAsnIleLysAspPheSerAsnPheI 58

1367 TCTATTGTGCTACAGTCTTCAGGGTTTTCAGTTATAGTAAGGTCTTTTGG 1318
     |||||         :::::::::      :::   ::::::::::::
  58 leTyrLeuLeuIleLeuTyrLys...AsnLysIleAsnAsnIleTyrAsn 73

1317 AGTTTGTACTTCAGGATTTAAATGTTTCCAAATATTTGTAAGAAAAATAT 1268
     :::   |||::::::|||                          ||||||
  74 AsnLysTyrTyrGluIle.......................LysAsnAs 82

1267 CTACATCGAATGATCCAGTTGGTAATTGTTCTTTC.........ATAAAT 1227
      |||||| |||      ::::::|||      :::         ||||||
  82 nTyrIle.AsnValPheLeuAsnAsnTyrTyrTyrLeuLysValIleAsn 98

1226 AAATGCCATATCCCTATACCATCTGCCCTATTATTATTTTGATTTTTATC 1177
     |||            |||   :::   |||
  99 Lys............IleGlnGlyIleLeu..................... 104

1176 ATGTATTGTTTTTCCTACCAATTCTAATAATTCTTTATTTTCGGGACTAA 1127
                                 ||||||:::|||::::::   :::|
 105 .........................AsnAsnAsnLeuTyrAsnLysIleA 113

1126 ATCCACCTCTGTTATGTTCTAATGAAAAACCTAATATGATTGTGATGTTA 1077
     |||||      :::   ||||||                    ::::::|||
 113 snPro......IleTyrSerAsn.................LeuPheLeu 121

1076 TATTTTAAACCAAAATATAAA........................... 1056
     :::|||:::   |||   |||
 122 PhePheAsnAsnLysIleLysIleLysTyrSerGlnLeuGlnGlnLeuIl 138

1055 ....TGGGATAAGTACATATCTACCATTATATCTGGATTTGTAGTTTCTA 1010
         :::      |||||||||:::   |||:::|||::::::   :::|
 138 eGlyTyrLysGlyTyrIleSerAsn...IleLysGlyMetIleTyrGluL 154

1009 AATTGTATGTT......TGAATAAATACAAGATCTATTTTCTCCTTTGCT 966
```

```
     ||      :::         ||||||      :::|||:::::::::
154 ysProValIleAsnAsnTyrIleAsnGluLeuAsnIleTyrGluTyrIle 170

966 .................................................. 966

171 LeuSerCysTyrGlySerLysLysGlyIleIleAspThrAlaLeuLysTh 187

965 ...............GTATTAAAAATATTAATAAAAGTTCCTGCTGATA 932
                   :::   |||   ||||||:::::::   ::::::
187 rAlaAspSerGlyTyrLeuThrLysArgLeuIleAsnIleThrSerAsnP 204

931 ACATCGCTGCTGCCCTA...TGTAATTCTTTATTTGTTTCTATTTGTTCA 885
     |||          |||   |||:::|||   |||:::   :::
204 heIleIleLysGluLeuAsnCysLysSerProPheIle...LeuLysTyr 219

884 GACAAAAATTTAGTGTTATATGGAGAT...................... 858
        |||:::   :::|||||||:::
220 IleLeuAsnMetAspIleTyrGlyAsnIleIleLeuProLeuAsnIleLe 236

857 ....TCTTCATCTTTACAGAAAGATGCAACTCCTGAAACGCATGATAATG 812
          :::   ||||||::::::
236 uArgPheLysIleLeuGlnAsnAsnIle..................... 245

811 CAGCATTTGATGAACCAGAAATTGAAATTAACTTTTCTTCCGGAATA... 765
                              :::|||:::::::::|||
246 .........................LeuAsnLeuAsnAsnGlyThrPhe 253

764 ............................GTTTTTCTTAACAAGTTAAT 745
                               :::|||||||||||||::
254 IleTyrThrLysAsnThrTyrIleThrLysTyrIleLeuAsnLysLeuLe 270

744 TAATTTAATATAAT.................AATTTGAAA......... 722
    :|||||| ||||||                 |||:::|||
270 uAsnLeu.TyrAsnArgArgAsnIleTyrLeuAsnIleLysSerValTyr 286

721 ......AATTTAATTCATTTAAGTTGTAAAACT................. 695
          |||:::   :::      |||:::|||
287 LeuCysAsnIleTyrAsnAsnIleCysAsnThrCysLeuAsnTyrLysGl 303

694 ....................TCCCATGTGGT..........TCCCAGT 677
                        :::|||:::            |||::::
303 nLeuTyrLysTyrAsnLeuGlyGlnHisIleGlyValIleSerSerGluA 320

676 CAATATCTACACCATCTAAATCAAAATCATTAACAAGCTTCAATATTTTA 627
    ::||||||   ||||||   :::      |||   :::|||:::
320 laIleSerGluProSerThrGlnMetValLeuArgThrPheHis...... 334

626 TCCACATAATCAATTTCTTTTTCTATATCTATCATATAGGTTTCTCCACC 577
                        :::   ||||||:::
335 .....................AlaSerSerIleLeu.............. 339

576 TAAGGATAAAAGAATAATTACATCTGGACGTACTTTTCTGATACGTCTAA 527
     |||||||||      :::::::::          :::   |||   |||
340 .LysAspLysPheAsnPheAsnLysTyrLeuIleTyrLysIleTyrLeuT 356

526 TTTCATTAAGCATCATACCA............TAGGTTTCATATTCTAAG 489
      :::|||:::|||                  :::::::::::|||
356 yrLysLeuAsnIleAsnLysIlePheLysLeuIleIleAsnPheLysLys 372

488 CCGTGTTTTCTTAATAGGAATCTTTGTCTTCCATTAAATGGTCTAGATAC 439
           ::::::   ||||||   :::   :::|||   :::
373 TyrIleAsnIleLysPheAsnLeuIlePheLeuMetAsnLysIleLeuTy 389

438 ATCATAT...AACATATTAATGCGAG..............CAAATGCAA 407
     :::|||   |||||||||::: |||               |||   |||
389 rAsnTyrAsnAsnIleLeuPhe.GluTyrLysTyrIleLeuGlnAsnGln 405

406 TATATAAAATTGACACCATTGGGTTTGAATCAATCATATGTTTTGCTCTA 357
    |||||||||   :::   :::      :::   |||      :::   ||
406 TyrIleLysCysAsnPheIleTyrAsnSerIleSerLysAsnPheLysTy 422

356 TCACCTTGACTGTTCCATGAACCATAGTATCCTGCTATAATACCTT.... 311
    |:::|||               ::::::|||::::::   ||||||
```

```
422 rAsnLeu..............AsnAsnIleIleIleLysTyrLeuAsnA 434

310 .....GTTTCCTTTTTT...TATACTCCTCTA.................. 287
          ::::::::::::   |||:::   :::
434 snValIleLysTyrTyrAsnTyrSerAsnIleGlnLeuLeuIleLysAsn 450

286 .........AAATTTGTCTTGGTGATTTTCTATCATTTTTATTTATATTT 246
             |||::::::|||      :::|||   :::|||||||||::
451 IleHisAsnLysTrpIleLeuTyrAsnIleTyrThrTyrTyrLeuTyrTy 467

245 TTATTATTTTTTGTAATTCTAGAGTTACACC................... 215
    :|||::::::      :::      |||:::
467 rTyrHisIleLysPheTyrAsnLeuTyrAsnLysGlyIleIleLeuAsnA 484

214 ............................CATCTCCGCAATAAGATCCAT 194
                                    |||   |||   ::::::
484 snAsnAsnAsnLysTyrAsnValIleTyrPheLeuIleAsnTyrPheAsn 500

193 ATTCTACAAAATTACTATTATTCGATTTAAGAT..............GT 159
    :::::::::|||||||||||||:::|||   :::             ::
501 LeuPheSerAsnTyrTyrTyrLysIleTyrAsnAsnAsnTyrAsnPheIl 517

158 GAAAAAGACTCATGTATTTCCGTTTGATGAGTTTTATTTTTA......TT 115
    :::::::::::::|||||||:::   |||:::|||||| |||      ||
517 eAsnSerAsnTyr.TyrPheLysLysMetAsnPheIle.LeuLysAsnPh 533

114 TTCCCGTATT...ATTCCCAATGAATTATTTATATTA......TTTTTTC 74
    |:::::::|||   |||   |||:::||||||   :::      :::|||
533 eAsnAsnIleGlnIleLeuAsnLysLeuPheTyrValAsnAsnIlePheI 550

 73 CTTTCAAGGTTCTGGAATTTGCAGAATACAAGATGGATACTATAATTAAA 24
      :::         |||         :::         :::||||||:::
550 leTyrTyrLysTyrGluLysLysLeuPheIleTyrLeuAsnIleIleAsn 566

 23 AATATTGATATTTTAAAATTC 3
    ||||||   |||   |||:::
567 AsnIleIleIleLysLysTyr 573
```

```
seq_name: pir2:T43916

seq_documentation_block:
chitinase A [imported] - Pyrococcus kodakaraensis
C;Species: Pyrococcus kodakaraensis
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T43916
R;Tanaka, T.; Fujiwara, S.; Nishikori, S.; Fukui, T.; Takagi, M.; Imanaka, T.
Appl. Environ. Microbiol. 65, 5338-5344, 1999
A;Title: A unique chitinase with dual active sites and triple substrate binding sites
A;Reference number: Z22722; MUID:20049967
A;Accession: T43916
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1215 <TAN>
A;Cross-references: EMBL:AB024740; NID:g6580044; PIDN:BAA88380.1; PID:g6580045
C;Genetics:
A;Gene: chiA


alignment_scores:
            Quality:  148.00            Length:    647
              Ratio:  0.576               Gaps:     28
 Percent Similarity:  39.722   Percent Identity: 17.311

alignment_block:
 US-09-579-383-2 x T43916   ..

 Align seg 1/1  to: T43916  from: 1  to: 1215

    160 CATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGG 209
        |||||||||:::|||::::::   :::                    ::
    139 HisLeuLysAlaAsnThrThrTyrTyrTyrGlyValValProValLeuAl 155

    210 AGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAA 259
```


155 aAspGly 157
260 ATGATAGAAAATCACCAAGACAAATT TTAGAG 291
158 ..SerArgGlySerProSerAsnValLeuAlaIleThrThrProLeuGlu 173
292 GAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATG 341
174 ProTyrArg ValIleValTyrTyrIleSerTr 184
342 GAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGG 391
184 pGlyArgTyrAla ArgLysPheTyrValSerAspIleProTrpG 199
392 TGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTA 441
199 luLysValThrHisValAsnTyrAlaPheLeuAspLeuLysGluAspGly 215
442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
216 ThrValAlaPhe 219
492 AGAATATGAAACCTATGGTATGATGCTTAAT GAAATTAGACGTA 535
220 ....TyrAspThrTyrAlaAspProLeuAsnLeuGluAlaMetLysGluT 235
536 TCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAA 585
235 yrLysArgLysTyrProAlaValLysValLeuIleSerValGlyGlyTrp 251
586 ACC TATATGATAGATATAGAAAAAGAAAT 614
252 ThrLeuSerLysTyrPheSerValValAlaAlaAspProAlaLysArgGl 268
615 TGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATG 664
268 nArgPheAlaGluThrAlaIleGluIleLeuArgLysTyrAsnLeuAspG 285
665 GTGTAGATATTGACTGGGAA CCACATGGGAAGTTTTACAAC 705
285 lyIleAspIleAspTrpGluTyrProGlyGlyGlyGlyMetAlaGlyAsn 301
706 TTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTT 755
302 TyrGluSerProAspAspGlyLysAsnPheValLeuLeuLeuLysAspLe 318
756 AAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAA 805
318 uArgGluAlaLeuAspLysAla 325
806 ATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAA 855
326 AlaLysGluAsp 329
856 GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAA 905
330 HisLysAspTyr 333
906 AGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATA 955
334 LeuLeuThrAlaAlaThrProAlaAspP 343
956 TTTTTAATACA GCAAAGGAGAAAATAGAT 984
343 roValLysAlaGlyArgIleAspTrpValGluAlaSerLysTyrLeuAsp 359
985 CTTGTATTTATTCAAACATAC 1005
360 SerIleAsnIleMetThrTyrAspTyrHisGlyAlaTrpGluThrIleTh 376
1006 AATTTAGAAACTACAAATC 1024


376 rGlyHisLeuAlaProLeuTyrCysAspProAsnAlaProTyrThrAspG 393
1025 CAGATATA ATGGTAGATATGTACTTA 1050
393 luAsnValLysTyrHisPheCysValAsnTyrThrValGlnTrpTyrIle 409
1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
410 GlnHisVal ProAspLysThrLysIleThrValGlyLeuProPh 424
1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTAT 1150
424 e...TyrSerArgSerPheAlaAsnValProProGluAsnAsnGlyLeuT 440
1151 TAGAATTGGTAGGAAAAACA 1170
440 yrGlnProPheSerGlyThrProAlaGlyThrTrpGlyProAlaTyrGlu 456
1171 ATACATGATAAAAATCAAAATAA 1193
457 ThrTyrGlyValMetAspTyrTrpAspValAlaGluLysAsnGlnSerSe 473
1194 T 1194
473 rGluTyrGluTyrHisTrpAspProIleAlaGlnValAlaTrpLeuTyrS 490
1194 1194
490 erProSerLysArgIlePheIleThrPheAspAspProArgAlaIleGly 506
1195 AATAGGGCAGATGGTATAGGGATATG 1220
507 IleLysValAspTyrMetLeuLysAsnGlyLeuGlyGlyValMetIleTr 523
1221 GCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATA 1270
523 pGluIleThrAlaAspArgLysProGlyThrAsnAspHisProLeu.... 538
1271 TTTTTCTTACAAATATTTGGAAACATTTAAAT 1302
539 .....LeuAspThrValLeuGlnHisLeuGlyGluLysProProAlaTrp 553
1303 ...CCTGAAGTA CAAACTCCAAAAGACCTTACTAT 1334
554 IleProAspThrTyrTyrIleGlySerAsnIleProSerAsnIleThrVa 570
1335 AACTGAA AACCCTGAAGACT 1354
570 lProGluProThrProLeuProProSerAsnGluThrThrProGluAspA 587
1355 GTAGCACAATA 1365
587 snGlnThrAsnProAsnProSerGlnGlyAsnGluThrAsnProAsnPro 603
1366 .....GATGAATATGTTCCA GGACTCGTTATTCCAACCAT 1400
604 SerProGlyAsnGluThrThrProSerAspAsnGlnThrThrProSerTh 620
1401 AGGGATATATTACAAACACAATGAT GCTATATGGA 1435
620 rGlyAspPheValLysProGlySerLeuSerValLysValThrAspTrpG 637
1436 AAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGG 1485
637 lyAsnThrGluTyrAspValThrLeuAsnLeuGlyGlyThrTyrAspTrp 653
1486 GACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCA 1535
654 ...ValValLysVal LysLeuLysAspGlySerSerValSe 666
1536 TTATTATAACACTGACTATAAA GAAAGCTCTATTATTATATGGA 1579
666 rSerPheTrpSerAlaAsnLysAlaGluGluGlyGlyTyrValValPheT 683

```
    1580 AAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCT 1620
                                 |||::::::||||||
     683 hrProValSer.............TrpAsnArgGlyPro 691

seq_name: pir2:S60651

seq_documentation_block:
chitinase precursor - Serratia marcescens
C;Species: Serratia marcescens
C;Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
C;Accession: S60651
R;Brurberg, M.B.; Eijsink, V.G.H.; Nes, I.F.
FEMS Microbiol. Lett. 124, 399-404, 1994
A;Title: Characterization of a chitinase gene (chiA) from Serratia marcescens BJL200 and
A;Reference number: S60651; MUID:95154677
A;Accession: S60651
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-563 <BRU>
A;Cross-references: EMBL:Z36294; NID:g908884; PIDN:CAA85291.1; PID:g908885
C;Superfamily: Serratia marcescens chitinase


alignment_scores:
           Quality:  145.50            Length:    425
             Ratio:  0.731              Gaps:     20
 Percent Similarity:  46.824   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x S60651   ..

 Align seg 1/1  to: S60651  from: 1  to: 563

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||                 ::::::|||:::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnSerLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspTh 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   ::::::::::::      |||:::|||
      73 rGlyThrThrAlaLysValLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ...............................AATATAAATAAAAATGAT 264
                                        ::::::|||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA........................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysSerAlaSe 123

     267 ........................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuAlaP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
              |||:::|||      :::||||||:::   |||      ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::         ::::::::::::|||:::   :::|||:::      ||
     155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         :::      :::|||:::   |||      ::::::   :
     171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
         :::::   ::::::   |||               |||   :::   |||
     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

     439 ................................GTATCTAGACCATTTAA 455
                                        :::::::|||      ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
         :|||||||::   |||   :::      |||            :::||||:
     219 nGlyArgGluAspPheLysValSerIleHisAspProPheAlaAlaLeuG 236

     494 AATATGAAACCTATGGTATGATG.......................... 516
         ::          |||:::
     236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::::|||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC..................TATATGATAG 598
         |   |||:::||||||   |||                   :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
            :::      :::|||||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluProProGlyGl 319

     693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
         ||||            ||||||::::::   :::   ::::::   |||:
     319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

     737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
         ::   |||::::::   ||||||      :::   :::
     336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                        |||||||::   :::|||
     349 ........................LeuSerAlaGluThrGly..... 354

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                              |||:::   |||:
     355 ................................ArgLysTyrGluLeuT 360

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         :::::   |||::::::::::|||:::            :::|||
     360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                     :::|||:::|||:::::::::::||
     375 ......................TyrAsnValAlaGlnAsnSerMetAs 383

     984 TCTTGTATTTATTCAAACATACAAT 1008
         |   :::|||:::::::::|||:::
     383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: pir2:A38221

seq_documentation_block:
chitinase (EC 3.2.1.14) MF1 - nematode (Brugia malayi)
C;Species: Brugia malayi
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: A38221
R;Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Piessens, W.F.; Perler, F.B.
```

```
Proc. Natl. Acad. Sci. U.S.A. 89, 1548-1552, 1992
A;Title: Transmission-blocking antibodies recognize microfilarial chitinase in brugian l
A;Reference number: A38221; MUID:92179220
A;Accession: A38221
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid; protein
A;Residues: 1-504 <FUH>
A;Cross-references: GB:M73689; NID:g156063; PIDN:AAA27854.1; PID:g156064
A;Note: sequence extracted from NCBI backbone (NCBIP:85345)
C;Keywords: glycosidase; hydrolase; polysaccharide degradation


alignment_scores:
           Quality:  145.00            Length:    445
             Ratio:  0.687              Gaps:     26
 Percent Similarity:  47.416   Percent Identity: 22.247

alignment_block:
 US-09-579-383-2 x A38221   ..

 Align seg 1/1  to: A38221  from: 1  to: 504

     313 GGTATTATAGCAGGA...TACTATGGTTCATGG.....AACAG 347
         |||   :::   |||   ||||||   :::|||          :::::
      22 GlyTyrValArgGlyCysTyrTyrThrAsnTrpAlaGlnTyrArgAspGl 38

     348 TCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAA 397
         ::::|||                  |||   ::::::      :::   |
      38 yGluGlyLysPheLeuProGlyAsnIleProAsnGlyLeuCysThrHisI 55

     398 TTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGA 447
         ||||||||   |||||||||:::::::::   |||   |||   |||:::
      55 leLeuTyr...AlaPheAlaLysValAspGluLeuGlyAsp...SerLys 69

     448 CCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATA 497
         |||||||:::   :::::::::                        :::
      70 ProPheGluTrpAsnAspGlu.....................AspTh 78

     498 TGAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTAC 547
          |||            |||   :::   :::   :::::::|||:::::::
      78 rGluTrpSerLysGlyMetTyrSerAlaValThrLysLeuArgGluThrA 95

     548 GTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTAT..... 591
         ::|||   :::   :::|||||||||   ||||||   ::::::
      95 snProGlyLeuLysValLeuLeuSerTyrGlyGlyTyrAsnPheGlySer 111

     592 ......ATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAATATT 635
               :::     |||   |||:::   :::   ::::::::::::::
     112 AlaIlePheThrGlyIleAlaLysSerAlaGlnLysThrGluArgPheIl 128

     636 GAAGCTTGTTAATGATTTT.....GATTTAGATGGTGTAGATA 673
         :|||   :::      |||           :::::||||||   |||:
     128 eLysSerAlaIleAlaPheLeuArgLysAsnAsnPheAspGlyPheAspL 145

     674 TTGACTGGGAA...CCACATGGGAAGTTTTACAACTTAAATGAATTAAAT 720
         ::|||||||||   |||   |||
     145 euAspTrpGluTyrProValGly.......................... 152

     721 TTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCC 770
            ::::::   :::   ||||||::::::   ::::::   ::::::
     153 ValAlaGluGluHisAlaLysLeuValGluAlaMetLysThrAlaPheVa 169

     771 GGAAGAAAAGTTAATTTCAATTTCTGGTTCATCA.....AATG 808
         ||||||         :::   ||||||::::::         :::|
     169 lGluGlu......AlaLysThrSerGlyLysGlnArgLeuLeuLeuThrA 184

     809 CTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAA 858
         |||||:::|||:::   :::|||:::   :::          :::   |||
     184 laAlaValSerAlaGlyLysGlyThrIleAspGlySerTyrAsnValGlu 200

     859 TCT.........CCATATAACACTAAATTTTTG.............. 882
         |||            ::::::      ||||||
     201 SerLeuGlyLysAsnPheAspLeuLeuPheLeuMetSerTyrAspLeuHi 217

     883 ..........TCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAG 922
                  :::::::::::::::   :::   :::||||||   :::
     217 sGlySerTrpGluLysAsnValAspLeuHisGlyLysLeuHisProThrL 234

     923 CAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATACA...... 966
           :::   :::|||:::          |||:::||||||||||||
     234 ysGlyGluValSerGly.........IleGlyIlePheAsnThrGluPhe 247

     967 ............................GCAAAGGAGAAAATAGATCT 986
                                     ||||||||||||   ::
     248 AlaAlaAspTyrTrpAlaSerLysGlyMetProLysGluLysIleIleIl 264

     987 T.........GTATTTATTCAAACATACAATTTAGAAACT.......... 1017
         :         ::::::   |||   ::::::|||::::::
     264 eGlyIleProMetTyrAlaGlnGlyTrpThrLeuAspAsnProSerGluT 281

    1018 ........................................ACAAAT 1023
                                                 ||||||
     281 hrAlaIleGlyAlaAlaAlaSerArgProSerSerAlaSerLysThrAsn 297

    1024 CCAGATATAATGGTAGATATGTAC.........TTATCCCATTTATATTT 1064
         |||          :::      |||           :::::::|||
     298 ProAlaGlyGlyThrAlaSerTyrTrpGluIleCysLysTyrLeuLysGl 314

    1065 TGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAACATA 1114
          |||   |||   ::::::                              |||:
     314 uGlyGlyLysGluThrVal.........................HisG 322

    1115 ACAGAGGT.............................GGATTTAGTCCC 1134
         :::::|||                                  ::::::
     322 lnGluGlyValGlyAlaTyrMetValLysGlyAspGlnTrpTyrGlyTyr 338

    1135 GAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAA 1184
         :::|||:::|||   :::::::::      |||   :::   :::|||::
     339 AspAsnGluGluThrIleArgIleLysMetLysTrpLeuLysGluLysGl 355

    1185 TCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTA....... 1227
         :   :::               |||      ||||||   |||
     355 yTyrGly..............GlyAlaPheIleTrpAlaLeuAspPheA 367

    1228 .....TTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATATT 1272
              |||      ::::::         |||   :::   :::
     367 spAspPheThrGlyLysSerCysGlyLysGlyProTyrProLeu...... 381

    1273 TTTCTTACAAATATTTGGAAACATTTA..............AATCCTGA 1307
           |||:::   |||   ::::::|||              ||||||||
     382 ...LeuAsnAlaIleSerSerGluLeuGluGlyGluSerGluAsnProGl 397

    1308 AGTACAAACTCCAAAA...GACCTTACTATAACTGAA............. 1341
         |:::   |||   :::   ::::::|||   ||||||
     397 uIleThrThrGluGluProSerIleThrGluThrGluAlaTyrGluThrA 414

    1342 ..AACCCTGAAGACTGTAGCACAATAGATGAATAT 1374
           :::   |||:::   |||       :::   |||
     414 spGluThrGluGluThrSerGluThrGluAlaTyr 425

seq_name: pir2:H71611

seq_documentation_block:
probable secreted protein PFB0565w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: H71611
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: H71611
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1817 <GAR>
```

```
A;Cross-references: GB:AE001403; GB:AE001362; NID:g3845216; PIDN:AAC71902.1; PID:g384521
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0565w


alignment_scores:
           Quality:  144.00            Length:    558
             Ratio:  0.514              Gaps:     31
 Percent Similarity:  50.179   Percent Identity: 20.968

alignment_block:
 US-09-579-383-2 x H71611   ..

 Align seg 1/1  to: H71611  from: 1  to: 1817

      64 ACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGA 113
         |||   :::||||||:::::::::|||:::
     355 ThrLysGlnGlyLysHisAspMetAsnAsp.................Ty 365

     114 AAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATC 163
          |||   ||||||   ::::::   :::|||:::   ::::::::::::
     365 rAsnMetAsnLysLysAsnAsnMetAspIleAsnIleThrIleAsnAsnA 382

     164 TTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT 213
           ::::::||||||:::|||   :::             |||      :::
     382 snAsnAsnAsnAsnAsnAsnLysIle............TyrAsnAspAsn 394

     214 GGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAA..... 258
             |||       :::   :::|||:::   ||||||::::::
     395 ThrLeuAsn......ValTyrAsnAsnSerTyrAsnIleHisSerAsnHi 409

     259 .......AATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAA 301
               ||||||:::::::::      |||:::||||||::::::   |
     409 sLeuMetAsnAspLysArgLysAsnAlaGlnValLeuGluLysHisLeuL 426

     302 AAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAA 351
         ||            :::::::::      :::      :::|||   :::
     426 ys............MetLeuCysAspAsnPhe......PheAsnLeuGlu 436

     352 GGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTT 401
                 :::::::::::::|||   :::|||      ::::::
     437 GluPheTyrSerSerAsnIleIleIleAsnAsnMetAspIleGluTyrTh 453

     402 ATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCAT 451
          |||      |||            :::||||||:::      :::   |
     453 rTyrAspTyrPhe............IleLeuTyrGlu......LysCysP 464

     452 TTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAA 501
         ||         ::::|||::::::                 ::::::|||
     464 heLeuProIleGluArgIleVal............HisValAsnTyrMet 476

     502 ACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCC 551
            |||   :::      |||:::   |||:::   :::|||:::|||
     477 LysTyr...LeuTyrLysAsnAsnGluArgLysLysAsnLysIleArgLy 492

     552 AGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATA 601
             :::|||   ||||||:::                |||      ||||
     492 sPheLeuIleThrLeuLeuGlu..............TyrSerArgAspI 504

     602 TAGAAAAAGAAATTGATTATGTGGATAAAATATTG...AAGCTTGTTAAT 648
         ||:::   ::::::         :::   ||||||   |||      |||
     504 leAsnPheHisPhePheIlePheAsnLeuIleLeuTyrLysCysLysAsn 520

     649 GATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTT 698
         :::|||      :::      ::::::      :::   :::      ||
     521 GluPheProCysSerIlePheGluLeuHisIleSerGlnTyrLeuTyrPh 537

     699 TTAC...AACTTAAATGAATTAAATTTTTCAAATTATTATATT....... 738
         |:::   :::|||||||||||||||||:::::::::   ||||||
     537 ePheValLysLeuAsnGluLeuAsnIleLysAspAlaTyrIleTyrTyrP 554

     739 .....................AAATTAATTAACTTGTTAAGAAAAACT 765

                                        :::|||   :::   ||||||:::
     554 heAsnAsnPheLysTyrGlnAspMetIleIleTyrPheSerArgLysAla 570

     766 ATT...CCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGC 812
         :::   |||   |||        :::::::::   :::::::::   ::
     571 PheTyrProTrpGlu......ThrAsnValGluGlnGlnLysLysGlnTh 585

     813 ATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTC 862
         :||||||   :::
     585 rLeuSerTyrIle..................................... 589

     863 CATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTA 912
           ||||||   |||             |||:::   ||||||:::
     590 ..TyrAsnAspLys..............IleLysLysAsnLysLys... 599

     913 CATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAAT........ 954
         :::::::::          ::::::   :::|||::::::|||
     600 AsnAsnSerTyrTyrGluMetAsnAsnAsnThrTyrMetAsnGluHisGl 616

     955 ......................ATTTTTAATACAGCAAAGGAGAAAATAG 982
                                    :::|||      :::::::::::::
     616 yTyrThrAspIleGluAsnGluArgLeuAsnLysLysAsnLysArgLeuA 633

     983 ATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAATCCAGATATA 1032
         :::::                  :::      :::|||      ||||||
     633 snVal...............ArgGlyArgThrAsnThrLeuAspAspIle 644

    1033 ATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATATAACAT 1082
         :::|||:::      :::|||              |||||||||
     645 IleValSer.........AspHisGlyAsnSerTyrAspLysTyrAsnTh 658

    1083 CACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTC 1132
          :::                    :::|||||||||
     658 rSer.....................LysHisAsnArg............. 663

    1133 CCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAA 1182
               :::|||:::   ::::::   :::   |||      ::::::|||
     664 .....ArgLysAsnHisIleAsnGluMetLysLysLysGlnAsnAsnLys 678

    1183 AATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTTAT 1232
         ::::::|||:::      :::||||||      :::      :::
     679 LysLysAsnThrLeuPheValAspGlyLysAspMetGluGlyIleGlyLy 695

    1233 GAAAGAACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAA 1282
          :::::::::      :::   :::::::::   :::||||||   :::|
     695 sGluLysGluLysGluAsnLysAsnMetAsnAsnAsnIlePheTyrAsnA 712

    1283 AT.......................ATTTGGAAACATTTAAATCCTGAA 1308
         ||                        ::::::::::::|||   :::
     712 snSerTyrSerAsnIleAsnAsnSerSerTyrSerAsnIleAsnAsnAsp 728

    1309 GTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAG 1358
         :::   :::      :::::::|||   ::::::|||   :::      ||
     729 IleTyrSerValAspAsnMetThrSerValAsnAsnThrLysTyrValSe 745

    1359 CACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATAT 1408
         |   :::   :::|||:::      |||:::         ::::::   |
     745 rGlyValProSerTyrAlaHisValLeuIleAsnLysGlnValAsnGluT 762

    1409 ATTACAAA............CACAATGATGCTATATGGAAAACTAGATCT 1446
         ||||||:::            :::|||:::   :::   |||      |||
     762 yrTyrGlnGlyLeuProAsnTyrAsnAsnMetMetIleLys...GlySer 777

    1447 TATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAA 1496
         :::   |||:::            ::::::|||
     778 HisIleIleAsnGluLeuProLysAsnAsnTyr................. 788

    1497 AGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCATTAT....... 1539
              ||||||:::          |||:::      ::::|||
     789 ....IleTyrGluAsnAsnTyrIleGlyGln......AsnTyrLeuMetT 802

    1540 ..........TATAACACTGACTATAAAGAAAGCTCTATTATTATATGG 1578
                   ||||||   :::   |||:::            |||:::
```

```
     802 hrAsnProLeuTyrAsnLysGluThrLysAspIlePheTyrThrIleTyr 818

    1579 AAAGGGGAACCATATTTAATTAAA 1602
         |||         ||||||:::|||
     819 Lys.........TyrLeuPheLys 823

seq_name: pir2:T24076

seq_documentation_block:
hypothetical protein R09D1.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24076
R;Matthews, P.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z19838
A;Accession: T24076
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-444 <WIL>
A;Cross-references: EMBL:Z70035; PIDN:CAA93865.1; GSPDB:GN00020; CESP:R09D1.5
A;Experimental source: clone R09D1
C;Genetics:
A;Gene: CESP:R09D1.5
A;Map position: 2
A;Introns: 79/1


alignment_scores:
            Quality:  143.00              Length:    470
              Ratio:  0.684                Gaps:     17
 Percent Similarity:  44.468   Percent Identity: 19.149

alignment_block:
 US-09-579-383-2 x T24076   ..

 Align seg 1/1  to: T24076  from: 1  to: 444

      89 ATAATTCATTGGGAATAATACGGGAAAATA...AAAATAAAACTCATCAA 135
         :::||||||:::      |||   |||:::   :::::::::
      19 ValIleHisPheIleAspTyrGlnLysLeuAlaGluLeuArgLysGluVa 35

     136 ACGGAAATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTT 185
          ::::::         :::      ||||||:::|||   |||:::::|
      35 lHisSerGluLysArgIleProSerIleLeuThrArgThrIleAlaValL 52

     186 TGTAGAATATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATT.... 231
         ||      :::   |||:::|||   :::::::::      |||
      52 euPheValPheAlaLeuLeuAlaLeuValSerIleGlySerGluGlyVal 68

     232 ......ACAAAAAATAATAAAAATATAAATAAAAATGATAGAAAATCACC 275
               :::|||:::::::::::::::            ||||||||
      69 ProGlnLeuArgAsnSerArgAspLeuThr...........LysSerPr 81

     276 AAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATTATAGCAG 325
         |                  |||||||||           ::::::|
      81 o.................CysLysLysArg...........ValValG 88

     326 GATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACATATGATT 375
         ||||||||::::::|||::::::                  |||
      88 lyTyrTyrSerGluTrpGluGlyThr.................GluIle 98

     376 GATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTAA 425
            :::   :::   ::::::   :::   :::||||||   |||::
      99 ThrArgSerGlnLeuGlyLysLeuThrHisAlaValPheAlaPheIleHi 115

     426 TATGTTATATGATGTATCTAGACCATTT......AATGGAAGACAAAGAT 469
         :|||      :::   :::      |||      |||   ::::::||||
     115 sMetAspSerGluGlyLysLeuGlnPheLysThrAsnGlnLysGluArgP 132

     470 TCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGATGCTT 519
         ||        |||
     132 heGluLysLeuLys.................................... 136

     520 AATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTT 569
                    :::::::::::::::   |||:::   ::::::::
     137 .........ThrAlaValLysAsnAlaAsnSerAspThrLysValMetIl 150

     570 ATCCTTAGGTGGAGAA.....................ACCTATATGATAG 598
         :|||:::||||||:::                    :::   :::   |
     150 eSerIleGlyGlyAspHisAsnSerGluAsnPheGlySerValLeuSerA 167

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||   ||||||:::      ::::::|||:::|||   ::::::::::::
     167 spSerGluLysLysSerMetPheIleAspSerIleAlaArgPheIleArg 183

     649 GATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTT 698
         :::      :::   ||||||||||||   |||            |||::
     184 GlnHisLysIleHisGlyValAspIleTyrTrp............LysTr 196

     699 TTACAACTTAAATGAATTAAATTTTTCAAATTAT......TATATTAAAT 742
         :   :::   :::|||   :::      ::::::      ::::::|||
     196 pLeuGlyAsnSerGluThrGluHisHisAspPheProSerPheLeuLysA 213

     743 TAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATT 792
           ::::::         ||||||:::   ::::::::::::|||||||||
     213 spLeuLysGluLysLeuLysThrValArgAspAspSerIleIleSerIle 229

     792 ..................................................... 792

     230 ValAlaProGlnAlaLysMetAspArgArgHisAspGlyTyrLysPheAs 246

     792 ..................................................... 792

     246 pAspPheMetGluTyrIleAspPheValAsnValPheSerMetAspTyrT 263

     793 ............................TCTGGTTCATCAAATGCT 810
                                     :::|||   |||
     263 yrGlyProTrpProAsnGlnTrpGlyThrProThrGlyProSerAlaPro 279

     811 GCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATC 860
               :::   ::::::         |||      |||:::
     280 LeuTyrGlyGlyIleGlyValLysLysHisPheAsnValAspSerThrMe 296

     861 TCCATATAACACT..................................... 873
             |||   |||
     296 tLysTyrTyrThrCysMetThrGluAspProSerLysPheAsnMetValI 313

     874 ......................AAATTTTGTCTGAACAAATAGAAACA 900
                                |||   ::::::|||   |||::::::
     313 leProPheTyrValArgLeuTrpLysAsnValLysGluProIleSerSer 329

     901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCA............GC 938
         :::   |||:::   |||   |||   :::   :::            ::
     330 GlyThrGluValPheArgArgAlaAspLeuLysAsnGlyAlaAlaValGl 346

     939 AGGAACTTTTATTAATATTTTTATACAGCAAAGGAGAAAATAGATCTTG 988
         ::::::::::::   :::::::::      |||      :::|||:
     346 yAsnSerTyrMetSerArgTrpThrValAspHisGluGlyTrpGluLeuT 363

     989 TA..........................TTTATTCAAACATACAATTTA 1011
         ::                                  ::::::|||
     363 hrProAlaLeuTrpAspAspValThrLysThrProTyrValTrpAsnGln 379

    1012 GAAACTACAAATCCAGATATAATGGTAGATATGTACTTATCCCATTTATA 1061
         ||||||   |||                                |||
     380 GluThrGlyAsn............................PheLeuTh 386

    1062 TTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAAC 1111
          |||      |||   :::|||          |||:::::::::::||||
     386 rPheGluAsnLysLysSerIleGluAlaLysLeuAlaTyrAlaIleGluH 403

    1112 ATAACAGAGGTGGATTT...................AGTCCCGAAAAT 1140
         |||||   ||||||                     :::   ||||||
     403 isAsnLeuGlyGlyValTrpIleHisLeuValAspLysAspAsnGluAsn 419
```

```
    1141 AAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAA 1190
         :::|||||||||||:::   |||:::::                :::::::
     420 GluGluLeuLeuArgAlaValAlaSerLysLysPheCysAlaSerGluSe 436

    1191 TAATAATAGG 1200
         ::::|||:::
     436 rGluAsnLys 439

seq_name: pir1:I30010

seq_documentation_block:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Leishmania tarentolae mitochondri
C;Species: mitochondrion Leishmania tarentolae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C;Accession: I30010
R;de la Cruz, V.F.; Neckelmann, N.; Simpson, L.
J. Biol. Chem. 259, 15136-15147, 1984
A;Title: Sequences of six genes and several open reading frames in the kinetoplast maxic
A;Reference number: A22848; MUID:85079995
A;Accession: I30010
A;Molecule type: DNA
A;Residues: 1-502 <DEL>
A;Cross-references: GB:M30010
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC6
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;


alignment_scores:
           Quality:  142.50              Length:    594
             Ratio:  0.582                Gaps:      34
 Percent Similarity:  41.246   Percent Identity: 20.707

alignment_block:
 US-09-579-383-2/rev x I30010   ..

 Align seg 1/1  to: I30010  from: 1  to: 502

    1743 TTCTTGTTCATATTGTTCCTCTACTTCTAGTGGTTTATGTGGATATTTTT 1694
         |||:::|||::::::|||:::   |||.:::   :::||||||   |||:
      16 PhePhePheMetPhePhePheMetPhe.GlyPheIleCysGlyIlePheP 32

    1693 TATTCCATTCTTCTATCCCTGGACATTTGGATGCATCTAGTTTTGTGTAT 1644
         :::::         ||| |||:::::|||:::   :::||||||||::
      32 heIleGlyArgHisIle.LeuSerPheTrpLeuSerValValLeuCysIl 48

    1643 GACTCTAGTGCCT..................................... 1631
         :   ||||||
      48 ePheLeuValMetSerThrIlePheSerCysPheCysLeuSerIleCysI 65

    1630 ...................GACCTTCCGGAGGTCCTTGTTGCCACCATT 1601
                            ||| |||         ::::::      ||
      65 leTyrGlyTyrCysPheTyrAsp.PheCysLeuLeuIleMetLeuAspPh 81

    1600 TAATTAAATATGGTTCCCCTTTCCATATAATAATAGAGCTTTCTTTATAG 1551
         |      ||||||:::   |||      |||:::      :::|||||:
      81 eCysPheIleTrpLeuValPheTyrCysAsnGlyPheTyrIlePheIleL 98

    1550 TCAGTGTTATA...ATAATGGGCTGCTTTCCCATCGCATAT...TTTTTC 1507
         ::   :::|||   |||:::   ||||||   :::   |||   |||:::
      98 euTyrLeuIleAspIleValPheCysPheIleValPheTyrAlaPheTyr 114

    1506 ATAGCATACTTTGACCAAGTC......CCATTCATATCTGTC........ 1471
               |||||||||:::::         |||:::   :::
     115 TyrMetTyrPheAspMetLeuLeuAlaArgPhePheHisIlePheTrpTr 131

    1470 .TACACCAGGTGCATGAATTGAATAAGATCTAGTTTTCCATAT....... 1429
          ::::::   |||||||| .::      |||:::   :::
     131 pPheValLeuCysMetAsnPhePheIleLeuSerTyrAspPheLeuThrA 148

    1428 .....................AGCATCATTGTGTTTGTAATATATCCC 1402

                                  :::::::::   |||   :::
     148 laTyrCysGlyTrpGluLeuLeuGlyLeuPheSerPhePheLeuIleSer 164

    1401 TATGGTTGGAATAACGAGTCCTGGAACATATTCATCTATTGTGCTACAGT 1352
         |||   |||   :::               |||
     165 TyrPheTrpTyrArg..............PheTyrAlaLeuLysPheGl 176

    1351 CTTCAGGGTTTTCAGTTATAGTAAG.........GTCTTTTGGAGTTTGT 1311
          |||:::::::|||       ||||||          |||:::       |||
     176 yPheLysAlaPhePheIleSerLysValGlyAspValLeuLeuLeuLeuV 193

    1310 ACTTCAGGATTTAAATGTTTCCAAAT....................... 1285
           |||   |||   ::::::   |||
     193 alPheAlaIleThrPheLeuMetAsnGlyTyrCysValIleThrPheTyr 209

    1284 ............ATTTGTAAGAAAAATATCTACATCGAATGATCCAGTTG 1247
                      :::|||         |||:::|||            ::
     210 PheLeuThrPheLeuCysValAspTyrIlePheIleValPheIleValPh 226

    1246 GTAATTGTTC.........TTTCATAAATAAATG................ 1222
         :   |||:::          |||   |||
     226 eLeuLeuLeuThrCysGlyPheThrLysSerThrGlnPheGlyLeuHisI 243

    1221 ..........................CCATATCCCTATACC......... 1207
                                     |||   |||   :::
     243 leTrpLeuProAspAlaMetGluGlyProIleProValSerAlaLeuIle 259

    1206 ..................ATCTGCCCTATTATTATTTTGATTTTTATCAT 1175
                           :::|||   ||||||::::::   |||:::::
     260 HisAlaAlaThrLeuValValCysGlyIleIleLeuIleSerPheValPh 276

    1174 GTATTGTTTTTCCTACCAATTCTAATAATTCTTT...ATTTTCGGGACTA 1128
         ::::||||||::::::   |||      |||:::   :::::::|||   :
     276 eTrpCysPheAspPheTrpPheCysTyrPheTyrSerLeuIleGlyTrpS 293

    1127 AATCC...ACCTCTGTTATGTTCTAATGAAAAACCTAATATGATTGTGAT 1081
         ::|||         ||||||:::         :::   |||:::   |||
     293 erSerLeuIleLeuValMetMetSerLeuCysValPheTyrAsnPheAsp 309

    1080 GTTATATTTTAAACCAAAATATAAATGGGATAAGTACATATCTACCATTA 1031
         |||                                          :::||
     310 Val.....................................LysArgTy 313

    1030 TATCTGGATTTGTAGTTTCTAAATTGTATGTTTGAATAAATACAAGATCT 981
         |:::         :::|||      :::|||      ::::::::::::
     313 rVal.........AlaPheSerThrIleCysGlnIleSerPheSerMetP 327

     980 ATTTTCTCCTTTGCTGTATTAAAAATATTAATAAAAGTTCCTGCTGATAA 931
         ::         ||||||::::::                :::|||
     327 he.........CysCysLeuSerLeuAspLeuTyrValGlyCys...... 338

     930 CATCGCTGCTGCCCTATGTAATTCTTTATTTGTTTCTATTTGTTCAGACA 881
                               :::|||:::|||::::::::::|||   ::
     339 ....................LeuPhePheCysTyrHisMetPheTyrLy 348

     880 AAAA......TTTAGTGTT......ATATGGAGATTCTTCATCTTT.... 847
         :          |||   |||      :::|||      |||:::|||
     348 sAlaThrLeuPheIleValLeuGlyValTrpIleHisPhePhePheGlyL 365

     846 .....ACAGAAAGATGCAACTCCTGAAACGCATGATAATGCAGCATTTGA 802
                |||||
     365 euGlnAspValArgCys.............................. 370

     801 TGAACCAGAAATTGAAATTAACTTTTCTTCCGGAATAGTTTTTCTTAACA 752
                                  :::|||   :::   |||
     371 .......................TyrPhePheThrTyrPheCysGlyCy 379

     751 AGTTAATTAATTTAATATAATAATTTGAAAAATTTAATTCATTTAAGTTG 702
          :::            ::::::|||            |||      :::
     379 sIleLeuAlaArgMetLeuLeuIle............PheAlaIleLeuA 392

     701 TAAAACTTCCCATGTGGTTCCCAGTCAATATCTACACCATCTAAATCAAA 652
           :::      :::||||||      ::::::|||         :::
```

```
     392 snSerCysSerLeuTrpPheLeuCysGlyPheTyrCysLysAspLeuLeu 408

     651 ATC...ATTAACAAGCTTCAATATTTTATCCACATAATCAATTTC..... 610
         :::  .:::      |||      |||:::   |||::::::|||
     409 LeuCysLeuLeuMetLeuThrSerPhePhePheIleLeuGluPheLeuCy 425

     609 .......TTTTTCTATATCTATCATATAGGTTTCTCCACCTAAGGATAAA 567
                 :::|||:::|||:::
     425 sValCysLeuPhePheIlePhePhe..................... 433

     566 AGAATAATTACATCTGGACGTACTTTTCTGATACGTCTAATTTCATTAAG 517
               :::   |||::::::||| ||||||:::   ::::::
     434 .....ThrValIleTyrAsnTyr.PheLeuLeuPhePheLeu..... 445

     516 CATCATACCATAGGTTTCATATTCTAAGCCGTGTTTTCTTAATAGGAATC 467
                                        ||||||:::   :::   :
     446 ..........................CysPheValPheLysCysP 452

     466 TTTGTCTTCCATTAAATGGTCTAGATACATCATATAACATATTAATGCGA 417
         ::||||||           :::||||||   :::   :::
     452 heCysLeu............ValAspThrLeuPheLeuLeu..... 461

     416 GCAAATGCAATATATAAAATTGACACCATTGGGTTTGAATCAATCATATG 367
                                       |||:::         ||
     462 ..............................PheAspPheGluCysCy 467

     366 TTTTGCTCTATCACCTTGACTGTTCCATGAACCATAGTATCCTGCTATAA 317
         |                                             ::::
     467 s.........................................LeuV 469

     316 TACCTTGTTTCCTTTTTTTATACTCCTCTAAAATTTGTCTTGGTGATTTT 267
         ::   |||   :::   ||||||        :::   |||
     469 alTyrCysThrPheCysLeuTyrMetCysPheValLeuLeuPhePheVal 485

     266 CTATCATTTTTATTTATATTTTTA 243
         |||:::||||||::::::|||:::
     486 LeuAspPheLeuTyrValPheIle 493

seq_name: pir2:A25090

seq_documentation_block:
chitinase (EC 3.2.1.14) - Serratia marcescens
C;Species: Serratia marcescens
C;Date: 12-Feb-1988 #sequence_revision 12-Feb-1988 #text_change 10-Dec-1999
C;Accession: A25090
R;Jones, J.D.G.; Grady, K.L.; Suslow, T.V.; Bedbrook, J.R.
EMBO J. 5, 467-473, 1986
A;Title: Isolation and characterization of genes encoding two chitinase enzymes from Ser
A;Reference number: A25090
A;Accession: A25090
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-561 <JON>
C;Superfamily: Serratia marcescens chitinase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation


alignment_scores:
            Quality:  142.50              Length:    425
              Ratio:  0.720                Gaps:     20
 Percent Similarity:  46.588   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x A25090   ..

 Align seg 1/1  to: A25090  from: 1  to: 561

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :

      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||              :::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   :::   :::|||      |||:::|||
      73 aGlyThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ............................AATATAAATAAAAATGAT 264
                                     ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA..................................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 ..................................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuProP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
               |||:::|||      :::|||||:::   |||      ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::         ::::::::::::|||:::   :::|||:::      ||
     155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         :::      :::|||:::   |||      ::::::   :
     171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
         :::::   ::::::   |||               |||   :::   |||
     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

     439 ..............................GTATCTAGACCATTTAA 455
                                      ::::::|||      ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
         :||||||:::   |||   :::      |||            :::|||:
     219 nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaAlaLeuG 236

     494 AATATGAAACCTATGGTATGATG.......................... 516
         ::            |||:::
     236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::::::||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
         |   |||:::||||||   |||                  :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
            :::      :::||||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

     693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
         ||||            ||||||::::::   :::   ::::::   |||:
     319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336
```

```
 737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
     ::   |||::::::   ||||||      :::   :::
 336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln..... 348

 787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                               |||||||:::   :::|||
 349 .........................LeuSerAlaGluThrGly..... 354

 837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                        |||:::   |||:
 355 ................................ArgLysTyrGluLeuT 360

 884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
     :::::   |||:::::::::|||:::          :::|||
 360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374

 934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                             :::|||:::|||:::::::::::||
 375 ......................TyrAsnValAlaGlnAsnSerMetAs 383

 984 TCTTGTATTTATTCAAACATACAAT 1008
     |   :::|||:::::::::|||:::
 383 pHisIlePheLeuMetSerTyrAsp 391
```

seq_name: pir2:B71621

seq_documentation_block:
probable membrane associated protein PFB0190c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71621
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: B71621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2295 <GAR>
A;Cross-references: GB:AE001379; GB:AE001362; NID:g3845118; PIDN:AAC71827.1; PID:g384511
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0190c

alignment_scores:
Quality: 139.50 Length: 731
Ratio: 0.431 Gaps: 37
Percent Similarity: 44.323 Percent Identity: 18.741

alignment_block:
US-09-579-383-2 x B71621 ..

Align seg 1/1 to: B71621 from: 1 to: 2295

```
   1 ATGAATTTTAAAATATCAATATTTTTAATTATAGTATCCATC..... 42
     :::|||||||   :::|||:::|||||||:::|||:::::::
   8 LeuAsnPheIlePheSerPhePheLeuLeuIleIleLysCysAspGluSe 24

  43 ......................TTGTATTCTGCAAATTCCAGAACCTTGA 70
                           :::|||      :::::::::   ||||
  24 rValSerAsnGlyArgLysGluIleTyrPheAspAspAspGluLysLeuL 41

  71 AA..................GGAAAAAATAATATAAATAATTCATTGGGA 102
     ||                  :::::::|||||||||   ::::::|||
  41 ysLeuSerSerPhePheAspArgSerThrAsnIleAsnLeuAspValGly 57

 103 ......................ATAATACGGGAA......AATAAAAA 122
                           :::   ||||||      :::|||::
  58 GluAsnAspGluLeuSerSerTyrValProArgGluValAspGluLysLy 74

 123 TAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTTAAATCGA 172
     :|||   ::::::   :::|||   :::|||   ::::::   |||||||:
  74 sLysLysAsnLysLysAspIle...AspSerLysGluAsnSerLysSerG 90

 173 ATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGTAAC 222
     ::|||:::   :::                                |||
  90 lyAsnAsnIleTyr..............................Asn 95

 223 TCTAGAATTACAAAAAATAATAAAAATATAAATAAAAAT.......... 261
     :::      ||||||||||||:::::::::|||   |||
  96 LysAspAsnThrLysAsnAsnGluAspValAsnTyrAsnValValLeuLy 112

 262 .GATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAAC 310
      |||   ::::::   :::   |||:::   :::   :::|||||||::::
 112 sAspGlyArgAlaLysGluGlyIleIleThrAsp...GluLysArgArgS 128

 311 AAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGA 360
     :::::         |||         :::   |||::::::      :::
 128 erSerThrLysAspGlyLysAsnLysGluGlnAsnAsnAsnLysMetAsn 144

 361 GCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGC 410
     :::      :::   |||:::|||
 145 SerAspAspValHisAspAsnAsn..................... 152

 411 ATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGAA 460
                 |||   :::   |||::::::
 153 .............AsnAsnMetAsnAspIleAsn............... 159

 461 GACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGT 510
           |||:::::::::                          |||:::
 160 ........PheValValGlu.....................TyrAsn 165

 511 ATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAAT 560
     |||:::::::::      :::|||                  |||   ::
 166 LysMetIleAspAsnTyrAspLysIleLeu............AspGluLe 178

 561 TATTCTTTTATCCTTAGGTGGAGAAACCTAT............ATGATAG 598
     :||||||   |||::::::   :::::::|||           |||:::|
 178 uIleLeuLysSerIleAsnArgAsnAsnTyrAsnTyrPheAsnMetLeuA 195

 599 AT.....................ATAGAAAAAGAAATT...GATTATGTG 624
     ||                  ::::::||||||:::   |||   :::
 195 spGluTyrSerLeuGlnThrLysLeuAsnLysGluMetTyrAspSerLeu 211

 625 GATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATAT 674
     :::   :::::::::|||:::|||:::   :::                ||
 212 AsnTyrLeuIleArgLeuMetAsnAsnLysAsnSerArgLysTyrPheIl 228

 675 TGACTGGGAACCACATGGGAAGTTTTACAACTTA...AATGAATTAAATT 721
     |:::::::::   :::   |||         :::   |||:::::::|||
 228 eSerPheSerAsnAsnGluLysLysLysIleIleLysAsnAspMetAsn. 244

 722 TTTCAAATTATTATATTAAA......TTAATTAACTTGTTAAGAAAAACT 765
       :::|||   |||||||:::      :::::::::|||:::|||
 245 ..GluAsnIleTyrIleArgHisPheIleValSerLeuPheArg...Trp 259

 766 ATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATT 815
           :::   ||||||||||:::         :::::|||      ::
 260 TyrAsnAsnPheLysLeuIleGluThrCysPheAspLysAsnAsnPheIl 276

 816 ATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTCCAT 865
     :      :::                          ||||||::::::   |
 276 eTyrTyrIle.....................AspGluAsnLysIleT 285

 866 ATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTACAT 915
     ||:::   |||:::                      |||   :::|||
 285 yrSerTyrLysTyr...................AsnTyrLysLeu... 293

 916 AGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATAC 965
                 ::::::|||:::   :::|||:::      :::|||
 294 ...MetLeuAsnLeuPheSerSerGluAsnPheLeuTyrTyrIleAsnLe 309

 966 AGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATACAAT....... 1008
```

```
     :::|||   ::::::   |||   :::|||::::::|||||| 
 309 uSerLysPheSerLeu...LeuGluIleIleAspAsnTyrAsnLysTyrS 325

1009 ........TTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050
             :::::::::   :::   |||          :::|||||::: 
 325 erPheIleIleAsnAsnIleLysArgAspTyrProAsnAsnMetTyrVal 341

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
       :::   ::::::   :::   |||:::                  :: 
 342 CysGlnSerPheTyrAspPheIleTyrSer.................Ty 352

1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTA. 1149
     :   |||::::::|||:::      |||      ::::::|||   ||| 
 352 rPheLeuSerTyrAsnHis...HisPhePheAspLysHisLysTyrLeuI 368

1150 .....TTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAAT 1194
          :::::::::   :::::::::|||:::   |||:::||| 
 368 leAsnMetAspIleTrpAsnAsnSerIleGlnThrLysGlyGln...... 382

1195 AATAGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAA........ 1236
                  ||||||        |||:::   ||| 
 383 ..............IleGlyAsnHisLysLeuTyrLysLysLeuLysLy 394

1237 .............................................G 1237
                                                 | 
 394 sLeuAsnGluAsnLeuIleLeuTyrAsnTyrIleLysAsnAspAspSerG 411

1238 AACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAAATATT 1287
     ||::::::|||      ::::::::::::   ::::::::::::|||::: 
 411 luMetIleProTyrValThrLeuGluMetArgMetIlePheSerAsnPhe 427

1288 TGGAAA.......................................... 1293
        ::: 
 428 ThrAsnLeuLeuIleAspIleLeuAsnLysLeuTyrAsnIleAspTyrGl 444

1294 ........................CATTTAAATCCTGAAGTACAAACTC 1318
                             ::::::||||||:::   ::::::| 
 444 nAspAsnIleLysGlnGluAsnValAsnValAsnProGlnArgAspAlaP 461

1319 CAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAGCACAATAGAT 1368
     ||:::|||   :::   ::::::|||      |||   |||      ::: 
 461 roGlnAspTyrValHisAsnLysAsnAspValAspValSerLeuLysAsn 477

1369 GAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAACA 1418
               |||      |||                      :::|| 
 478 ValLysGluProLysLysVal.....................GluHi 486

1419 CAATGATGCTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTG 1468
     ||||   |||:::            :::|||:::            ||| 
 486 sAsnLysAlaMet.........SerAsnTyrGluThrAspGluArgGlyA 500

1469 TAGACAGATATGAATGGGACTTGGTCAAAGTATGCTATGAAAAAATA... 1515
            |||:::   :::      |||      :::|||||| 
 500 spMetIleTyrAsp...AsnThrAsnLysGluLysPheGluLysSerGlu 515

1515 ........................................................ 1515

 516 GlyThrPheAsnAsnIleSerGlyGlyGluAspThrPheLysAsnIleSe 532

1515 ........................................................ 1515

 532 rGlyGlyGluAspThrPheLysAsnIleSerGluGlyAspGlyGluValA 549

1516 ........................TGCGATGGGAAAGCAGCC...... 1533
                             ||||||:::::::||| 
 549 spGlyAspGlyGluGlyAspGlyAspGlyAspGlyGluGlyAlaAspAsp 565

1534 ..............CATTATTATAACACTGACTATAAAGAAAGCTCTAT 1568
                  |||      ::::::|||   |||||||||::: 
 566 SerSerValAspThrHisAsnAsnLysAsnAspGlyLysGluSerGluSe 582

1569 TATTATATGG...................................... 1578
        :::||| 
```

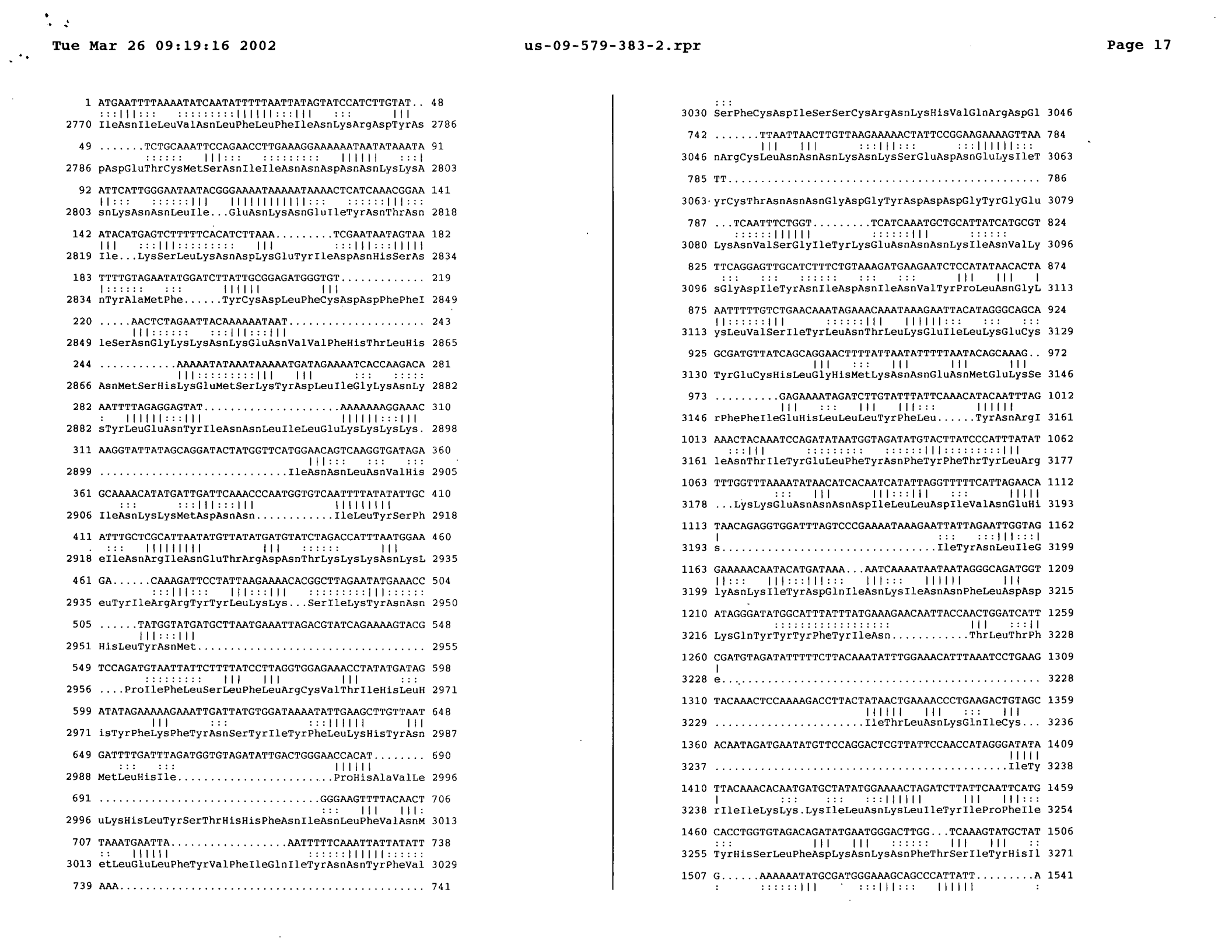
```
 582 rAspValTrpAsnLeuLeuMetAspSerTyrLysLysLeuAlaAsnAspG 599

1579 ..........AAAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGA 1617
                |||   :::   |||:::::::||| 
 599 luAsnPheLysLysTyrAsnLysTyrIleLeuLys.............. 610

1618 CCTCCGGAAGGTCAGGCACTAGAGTCATACACAAAACTAGATGCATCCAA 1667
                          |||::::::::::   ::::::::::::::|| 
 611 ..............AsnLeuAspLysPheLeuAsnMetSerSerGluLy 622

1668 ATGTCCAGGGATAGAAGAATGGAATAAAAAATATCCACATAAA 1710
     |         |||:::::::::::::::::||||||      ||| 
 622 sLysGluAspIleAsnSerTyrLysAsnLysTyrGluLeuLys 636
```

```
seq_name: pir2:T30933

seq_documentation_block:
chitinase (EC 3.2.1.14) A - Pseudoalteromonas sp. (strain S9)
C;Species: Pseudoalteromonas sp.
A;Variety: strain S9
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T30933
R;Techkarnjanaruk, S.; Goodman, A.E.
Microbiology 145, 925-934, 1999
A;Title: Multiple genes involved in chitin degradation from the marine bacterium Pseu
A;Reference number: Z20935; MUID:99235578
A;Accession: T30933
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1054 <TEC>
A;Cross-references: EMBL:AF007894; NID:g3928771; PID:g3928775; PIDN:AAC79665.1
C;Genetics:
A;Gene: chiA
C;Keywords: glycosidase; hydrolase


alignment_scores:
           Quality:  138.50            Length:    581
             Ratio:  0.584               Gaps:     29
 Percent Similarity:  40.792   Percent Identity: 19.105

alignment_block:
 US-09-579-383-2 x T30933   ..

 Align seg 1/1  to: T30933  from: 1  to: 1054

     319 ATAGCAGGATACTATGGTTCATGG...AACAGTCAAGGTGATAGAGCAAA 365
         :::   |||||||:::   ||||||   |||:::   :::   :::   :: 
     322 ValIleGlyTyrPheThrSerTrpArgAsnGlyAlaAsnGlyGlnProSe 338

     366 ACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATT......G 409
         ::::::::::|||:::   |||      :::|||   :::|||      | 
     338 rTyrLeuValAspAspIleProTrpAspLysIleThrHisIleAsnTyrA 355

     410 CATTTGCTCGCATT.................................. 423
         ||||||||::::::
     355 laPheAlaHisValAspAlaAsnAsnLysValSerIleGlyAspProSer 371

     424 .....................AATATGTTATAT.............. 435
                              ||||||   ::: 
     372 AlaAlaGlyAsnProAlaThrAsnMetGluTrpProGlyValAlaGlyAl 388

     436 .......GATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTATTAA 478
               |||   :::   |||:::::::|||::: 
     388 aGluMetAspProThrLeuProTyrLysGlyHis.............. 399

     479 GAAAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATT 528
                                                 :::|||   ::: 
     400 ..................................PheAsnLeuLeu 403

     529 AGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGG 578
         ::::::   :::|||::::::|||||||||      |||:::|||:::|| 
     404 AsnLysTyrLysLysLeuHisProAspValLysThrLeuIleSerValGl 420
```

```
  579 TGGA..................................................... 582
      ||||
  420 yGlyTrpAlaGluThrGlyGlyTyrPheAspAlaAsnGlyAsnArgValA 437

  583 ..............GAAACCTATATGATAGATATAGAAAAAGAAATTGAT 618
                       |||          :::   :::   :::::::::
  437 laSerGlyGlyPheTyrThrMetThrThrAsnAlaAspGlySerValAsn 453

  619 ..................TATGTGGATAAAATATTGAAGCTTGTTAATGA 650
                         ::::::   :::   ::::::::::::::::
  454 ThrAlaGlyIleAsnAlaPheAlaLysSerAlaValGluPheIleGluTh 470

  651 TTTTGATTTAGATGGTGTAGATATTGACTGGGAA...CCACATGGGAAGT 697
       :::   :::|||||||||||||||||||||:::|||   |||
  470 rTyrGlyPheAspGlyValAspIleAspTyrGluTyrPro..... 483

  698 TTTACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATT 747
                 :::::::::|||   |||::::::
  484 ...........SerSerMetAsnAspSerGlyHis.............. 491

  748 AACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGG 797
                           |||::::::      :::   ||||||::
  492 .....................ProAspAsp......PheProIleSerAs 499

  798 TTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTA 847
      ::::   :::|||:::|||::::::         |||
  499 nAlaArgArgAlaGlyLeuAsnAlaSerTyrGlnValLeu..... 512

  848 AAGATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAA 897
                              |||   |||   |||:::::::::
  513 ......................MetLysThrLeuArgGluGluLeuAsp 521

  898 ACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA.............. 933
            ||||||                 ||||||
  522 LysAlaGlyGluLeuAlaGlyLysHisTyrMetLeuThrIleAlaSerPr 538

  934 .TCAGCAGGAACTTTTATT.........AATATTTTTAATACAGCAAAGG 973
       |||:::|||   ::::::         :::   |||:::|||:::|||
  538 oSerSerGlyTyrLeuLeuArgGlyMetGluThrPheGlnThrThrLys. 554

  974 AGAAAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAAT 1023
        :::|||   |||   |||::::::|||:::|||:::   :::
  555 ..TyrLeuAspTyrValAsnIleMetSerTyrAspLeuHisGlyAlaTrp 570

 1024 CCAGAT..................................................... 1029
        |||
  571 AsnAspHisValGlyHisAsnAlaAlaLeuTyrAspThrGlyLeuAspSe 587

 1030 .......ATAATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAA 1072
              :::   ::::::|||   :::       :::   |||:::
  587 rGluLeuAlaGlnTrpAsnValTyrGlyThrLysGluPheGluGlyIleG 604

 1073 AATATAACATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGT 1122
        |||      |||            :::::::::   :::   ||||||
  604 lyTyrLeuAsnThr.........AspTrpAlaValThrTyrPheArgGly 617

 1123 GGATTTAGT..................................................... 1131
      |||:::|||
  618 GlyLeuSerAlaGlyArgIleAsnIleGlyValProTyrTyrThrArgGl 634

 1131 ..................................................... 1131

  634 yPheLysAspValSerGlyGlyGluAsnGlyLeuTrpGlyArgAlaAlaL 651

 1132 ..CCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGAT 1179
        |||::::::   :::       :::       |||       :::   :::
  651 euProAsnGlnAlaAspCysAlaLys......GlyThrGlyValGlyGlu 665

 1180 AAAAATCAAAATAATAATAGGGCAGATGGTATA...GGGATATGGCATTT 1226
      ||||||:::   :::|||   |||   ||||||   ::::::||||||
  666 LysAsnLysCysGlyAsnGlyAlaValGlyIleAspAsnMetTrpHisAs 682
```

```
 1227 ATTTATGAAA......GAACAATTACCAACTGGATCATTCGATGTAGATA 1270
        :::   :::       :::::::::|||:::||||||      :::
  682 pLeuAspGluAsnGlyAsnGluMetProAlaGlySerAsnProLeu...T 698

 1271 TTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAAGTACAAACTCCA 1320
           :::   |||:::   :::   :::   |||:::      :::
  698 rpHisValLysAsnLeuGlnAsnGlyIleLeuProSerTyrLeuAlaAla 714

 1321 AAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAGCACAATAGATGA 1370
             |||      :::::::::|||   |||
  715 TyrGlyLeuAspProAlaAsnAspProThrAsp..... 725

 1371 ATATGTTCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAACACA 1420
                                  :::|||   |||   :::|||
  726 .....................LysLeuValGlyThrTyrThrArgHisT 735

 1421 ATGATGCTATA..................TGGAAAACTAGATCTTATTCA 1452
        |||::::::                  |||:::::: |||
  735 yrAspSerValAlaValAlaProTrpLeuTrpAsnAla.AspLysAsnVa 751

 1453 ATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAAAGTAT. 1501
       |||:::   :::         :::       ||||||          |||
  751 lPheLeuSerIleGluAspGluGluSerMetGlyThrLysLeuAspTyrV 768

 1501 ..................................................... 1501

  768 alIleAsnLysGlyLeuGlyGlyIleMetPheTrpGluLeuAlaGlyAsp 784

 1502 ..............GCTATGAAAAAATATGCGATGGGAAAGCAGCCCAT 1536
                     |||   |||:::|||   ||||||:::
  785 PheAspTyrAspSerAlaLysLysGluTyrPheMetGlySer........ 798

 1537 TATTATAACACTGACTATAAAGAAAGCTCTATTATTATATGGAAAG.... 1582
             |||:::|||               |||   |||:::|||
  799 .......ThrMetThr............SerLeuAlaTyrAsnLysPheA 809

 1583 ........GGGAACCATATTTAATTAAATGGTGGCAACAAG 1615
              |||:::          :::|||   |||:::|||
  809 laGlnAlaGlySerLysTyrAspIleAsnLysGlyGluLys 822
```

```
seq_name: pir2:B71612

seq_documentation_block:
hypothetical protein PFB0555c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71612
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: B71612
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3973 <GAR>
A;Cross-references: GB:AE001402; GB:AE001362; NID:g3845213; PIDN:AAC71900.1; PID:g384
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0555c
C;Superfamily: Plasmodium falciparum hypothetical protein PFB0555c


alignment_scores:
            Quality:  138.00              Length:    675
              Ratio:  0.471                 Gaps:     34
 Percent Similarity:  43.407   Percent Identity: 20.889

alignment_block:
 US-09-579-383-2 x B71612   ..

 Align seg 1/1  to: B71612  from: 1  to: 3973
```



```
    3271 eAsnAsnAsnGlnTyrIleArgAsnLysAspHisIleIlePheCysSerL 3288

    1542 TAACACTGACTATAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCAT 1591
         ::   :::   |||   |||   |||   ::::::   |||
    3288 euIleValPheIleIleLysLeuLeuTyrValPheValLys..... 3301

    1592 ATTTAATTAAATGGTGGCAACAAGGACCTCCGGAAGGTCAGGCACTAGAG 1641
                      |||           ::::::|||:::      ::
    3302 .................AsnPheLysHisLysAsnValAsnPheAsnAs 3312

    1642 TCATACACAAAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATGGAA 1691
         :|||   :::|||      |||   ||||||
    3312 nHisGlnAsnAsnLysAsnHisGluAsnVal.................. 3322

    1692 TAAAAAATATCCACATAAACCAC 1714
              |||   :::::::::|||
    3323 .....AsnThrAsnValGlyHis 3328

seq_name: pir1:QQUTC5

seq_documentation_block:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Trypanosoma brucei mitochondrion
C;Species: mitochondrion Trypanosoma brucei
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-May-2000
C;Accession: A04519; D22845
R;Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.
Nucleic Acids Res. 12, 7327-7344, 1984
A;Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift conta
ondrial maxi-circle DNA.
A;Reference number: A93537; MUID:85037915
A;Accession: A04519
A;Molecule type: DNA
A;Residues: 1-590 <HEN>
A;Cross-references: GB:M94286; NID:g343546
A;Note: this translation is not annotated in GenBank entry TRBKPGEN, release 109.0
C;Comment: The DNA sequence is from a segment of the 20-kb maxicircle, which is believed
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC6
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;


alignment_scores:
            Quality:  136.50              Length:    661
              Ratio:  0.533                Gaps:     35
 Percent Similarity:  38.729   Percent Identity: 19.062

alignment_block:
 US-09-579-383-2/rev x QQUTC5   ..

 Align seg 1/1  to: QQUTC5  from: 1  to: 590

    1696 TTTTATTCCATTCTTCTATCCCTGGACATTTGGATGCATCTAGTTTTGTG 1647
         |||   |||:::   |||||||||   :::   :::   :::|||||||| .
      25 PheLeuSerPheTrpLeuSerLeuValMetIleIlePheIleValLeuCy 41

    1646 TATG.....................ACTCTAGTGCCTG............ 1630
         ||||                     :::::: ||||||
      41 sMetIlePheSerPheLeuMetValSerVal.CysLeuTyrGlyTyrTyr 57

    1629 ......ACCTTCCGGAGGTCCTTGTTGCCACCATTTAATTAAATATGGTT 1586
                |||          |||:::      |||      ||||||::
      58 TyrTyrAspPheCysLeuIleLeuMetLeuAspPheCysPheIleTrpLe 74

    1585 CCCCTTTCCATATAATAATAGAGCTTTCTTTATAGTCAGTGTTATA...A 1539
         :   :::      ::::::      :::||||||:::   :::|||   :
      74 uThrTyrValCysSerGlyPheTyrMetPheIleMetLeuLeuIleAsnM 91

    1538 TAATGGGCTGCTTTCCCATCGCATAT...TTTTTCATAGCATACTTTGAC 1492
         :::::   ||||||   :::   |||   |||:::      |||||||||
      91 etValPheCysPheIleValPheTyrAlaPheTyrTyrMetTyrPheAsp 107

    1491 CAA..............GTCCCATTCATATCTGTCTACACCAGGTGCAT 1457
         :::                  :::   ::::::   :::::::::   |||||
     108 MetLeuLeuGlyArgPheLeuIleIlePheTrpIlePheValValCysMe 124

    1456 GAATTGAATAAGATCTAGTTTTCCATAT..................... 1429
         ||||   :::      |||:::   :::
     124 tAsnLeuPheIleLeuSerTyrAspPheLeuThrAlaTyrCysGlyTrpG 141

    1428 ........AGCATCATTGTGTTTGTAATATATCCCTATGGTTGGAATAAC 1387
                 :::::::::   |||   :::      |||   |||
     141 luLeuLeuGlyLeuPheSerPhePheLeuIleSerTyrPheTrp...... 155

    1386 GAGTCCTGGAACATATTCATCTAT......................... 1363
              ::::::::::|||
     156 ......TyrArgPhePheAlaLeuLysPheGlyPheLysAlaPhePheIl 170

    1362 .......TGTGCTACAGTCTTCAGGGTTTTCAGTTATAGTAAGGTCTTTT 1320
                 :::   |||:::   :::|||::::::|||   :::|||
     170 eGlyLysIleGlyAspValLeuLeuIlePheAlaPheSerIleIlePheL 187

    1319 GGAGTTTGTACTTCAGGATTTAAATGTTTCCAAATATT.........TGT 1279
           |||      |||   :::   |||      :::          |||
     187 euSerAsnGlyPheCysMetThrThrPheTyrPheLeuAsnPhePheCys 203

    1278 AAGAAAAATATCTACATCGAATGATCCAGTTGGTAATTGTTC........ 1237
                     |||||||||   |||         |||
     204 MetAspTyrTyrTyrIleGluPheSerIleCysLeuLeuValGlyCysAl 220

    1236 .TTTCATAAATAAATG............................... 1222
          |||   |||
     220 aPheThrLysSerThrGlnPheGlyLeuHisIleTrpLeuProAspAlaM 237

    1221 ........CCATATCCCTATACC......................... 1207
                 |||   |||   :::
     237 etGluGlyProIleProValSerAlaLeuIleHisAlaAlaThrLeuVal 253

    1206 ATCTGCCCTATTATTATTTTGATTTTTATCATGTATTGTTTTTCCTACCA 1157
         :::|||   ||||||:::|||   |||:::   :::||||||:::::
     254 ValCysGlyIleIleLeuLeuSerPheValTyrTrpCysPheAspPheTr 270

    1156 ATTCTAATAATTCTTT...ATTTTCGGGACTAAATCC............. 1123
          |||      |||:::   :::::::|||   ::::::
     270 pPheSerTyrPheTyrAsnLeuIleGlyTrpSerThrLeuIleLeuIleL 287

    1122 .....ACCTCTGTTATGTTCTAATGAAAAACCTAATATGATTGTGATGTT 1078
               |||      :::|||                |||:::   ||||||
     287 euMetThrLeuCysValPhe..............TyrAsnPheAspVal 298

    1077 ATATTTTAAACCAAAATATAAATGGGATAAGTACATATCTACCATTATAT 1028
                                                  :::|||::
     299 .....................................LysArgTyrVa 302

    1027 CTGGATTTGTAGTTTCTAAATTGTATGTTTGAATAAATACAAGATCTATT 978
         :         :::|||      :::|||      ::::::::::::::
     302 l.........AlaPheSerThrIleCysGlnIleSerPheSerMetPheC 316

     977 TTCTCCTTTGCTGTATTAAAAATATTAATAAAAGTTCCTGCTGATAACAT 928
               ||||||           |||::::::
     316 ysCysLeuCysIleAspIleTyrIleGlySer.................. 326

     927 CGCTGCTGCCCTATGTAATTCTTTATTTGTTTCTATTTGTTCAGACAAAA 878
                              :::|||:::|||:::::::::|||   :::
     327 ..................LeuPhePheCysTyrHisMetPheTyrLysAl 337

     877 A.....TTTAGTGTT......ATATGG..................... 862
               |||   |||      ||||||
     337 aThrLeuPheIleValLeuGlyIleTrpIleHisIlePhePheGlyLeuG 354

     861 ..............AGATTCTTCATCTTTACAGAAAGATGCAACTCCTGA 826
                           ||||||:::::          |||  :::|||
     354 lnAspLeuArgCysTyrPhePheMetTyrPheCysGlyCys..ValLeuA 370

     825 AACGCATGATAATGCAGCATTTGATGAACCAGAAATTGAAATTAACTTTT 776
```

```
      |||:::::::::          :::   ::::::   ::::::   |||
370 laArgLeuLeuLeuIlePheAlaIleLeuAsnSerCysSerIleTrpPhe 386

775 CTTCCGGAATAGTTTTTCTTAACAAGTTAATTAATTTAATATAATAATTT 726
    |||
387 Leu..................................................... 387

725 GAAAAATTTAATTCATTTAAGTTGTAAAACTTCCCATGTGGTTCCCAGTC 676
                        |||
388 ......................CysGly...................... 389

675 AATATCTACACCATCTAAATCAAAATCATTAACAAGCTTCAATATTTTAT 626
                                                  ||||||
390 ..............................................PheTyr 391

625 CCACATAATCAATTTCTTTTTCTATATCTATCATATAGGTTTCTCCACCT 576
        ::::::::::|||   |||   :::         |||   :::
392 CysLysAspMetLeuLeuAlaLeuLeuMetLeuLeuSerPheTyrAsn.. 407

575 AAGGATAAAAGAATAATTACATCTGGACGTACTTTTCTGATACGTCTAAT 526
      |||   |||   |||   ::::::::::::|||   :::
408 ....IleIleGluPheLeuPheIleSerIleIlePheIlePhe..... 420

525 TTCATTAAGCATCATACCATAGGTTTCATATTCTAAGCCGTGTTTTCTTA 476
                          |||   ::::::   :::   ||||||
421 ......................PheThrMetIleTyrAsnTyrPheLeu 429

475 ATAGGAATCTTTGTCTTCCATTAAATGGTCTAGATACATCATATAACATA 426
    :::   ::::::::::|||                          :::::
430 LeuPheLeuMetPhe...........................ValPh 437

425 TTAATGCGAGCAAATGCAATATATAAAATTGACACCATTGGGTTTGAATC 376
    :   |||                            :::   |||   :
437 eLysCys..............................PheCysLeuValA 444

375 AATCATATGTTTTGCTCTATCACCTTGACTGTTCCATGAACCATAGTATC 326
    ::        ||||||:::      :::|||   :::   :::   :::
444 spCysLeuPheLeuLeuPheAspTyrGluCysCysLeuValTyrCysLeu 460

325 CTGCTATAATAC......CTTGTTTCCTTTTTTTATACTCCTCTAAAATT 282
    :::      |||      ::::::|||:::|||:::      :::   ||
461 IleSerLeuTyrMetCysIleLeuSerIlePhePheIle...IleAspPh 476

281 TGTCTTGGTG.....................ATTTTCTATCATTTTTATT 253
    ||||   :::                     :::|||:::   |||:::|
476 eValCysIlePheValPheSerSerTyrCysValPheTrpSerPhePheL 493

252 TATATTTTTATTATTTTTTGTAATTCTAGAGTTACACCCATCTCCGCAAT 203
    ||   ||||||   |||
493 euAsnPheTyrAsnPhe................................ 498

202 AAGATCCATATTCTACAAAATTACTATTATTCGATTTAAGATGTGAAAAA 153

498 ..................................................... 498

152 GACTCATGTATTTCCGTTTGATGAGTTTTATTTTTATTTTCCCGTATTAT 103
                  ||||||         |||         :::::::
499 ...............PheAspIleAlaIlePheValValPheLeuIleL 510

102 TCCCAATGAATTATTTATAT.................TATTTTTTC... 74
    ::   :::::::::||||||                 |||||||||
510 euSerValGlyPheLeuTyrTyrGlyCysLeuPhePheTyrPhePheAsn 526

 73 .........CTTTCAAGGTTCTGGAATTTG 53
            :::      ||||||::::::
527 IleAspCysIleMetLeuPheTrpArgIle 536
```

```
OM of: US-09-579-383-2 to: SwissProt_39:*   out_format : pfs

Date: Mar 21, 2002  4:54 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09579383/runat_20032002_151240_29459/app_query.fasta_1.1849
 -DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09579383_@CGN1_1_43 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-579-383-2
Query length: 1764
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36664827
Search time (sec): 79.320000



score_list:
Sequence              Strd Orig    ZScore   EScore Len   ! Documentation  ..
SwissProt_39:CHIT_BRUMA +  145.00  170.16  0.0166    504  ! P29030 brugia malayi. endoch
SwissProt_39:CHIA_SERMA +  143.50  167.48  0.0209    563  ! P07254 serratia marcescens.
SwissProt_39:NU5M_TRYBB -  136.50  158.80  0.0607    590  ! P04540 trypanosoma brucei br
SwissProt_39:KTXA_KLULA +  136.00  152.79  0.0676   1146  ! P09805 kluyveromyces lactis
SwissProt_39:CYSP_PLAFA +  131.50  153.17  0.1296    569  ! P25805 plasmodium falciparum
SwissProt_39:RPC1_PLAFA +  131.00  141.04  0.1493   2339  ! P27625 plasmodium falciparu
SwissProt_39:PTP1_DICDI +  129.50  151.51  0.1750    521  ! P34137 dictyostelium discoid
SwissProt_39:CHIT_NPVAC +  129.50  151.06  0.1754    551  ! P41684 autographa californic
SwissProt_39:Y832_METJA +  124.50  135.71  0.3957   1750  ! Q58242 methanococcus jannas
SwissProt_39:RPOB_PLAFA -  121.50  136.52  0.6090   1024  ! P21421 plasmodium falciparu
SwissProt_39:YXDM_BACSU +  121.00  140.00  0.6422    622  ! P42424 bacillus subtilis. hy
SwissProt_39:CHIT_NPVOP +  120.00  139.81  0.7434    550  ! O10363 orgyia pseudotsugata
SwissProt_39:HEMK_RICPR +  119.00  139.12  0.8631    518  ! Q9zcb3 rickettsia prowazekii
SwissProt_39:Y875_METJA +  116.50  133.16  1.28      748  ! Q58285 methanococcus jannasc
SwissProt_39:RPOP_AGABT +  116.50  130.00  1.31     1102  ! P33539 agaricus bitorquis.
SwissProt_39:IF2P_METJA +  116.50  129.61  1.31     1155  ! Q57710 methanococcus jannas
SwissProt_39:CHI1_COCIM +  116.00  137.14  1.35      427  ! P54196 coccidioides immitis.
SwissProt_39:NTP2_MSEPV +  116.00  132.91  1.38      717  ! Q9yw06 melanoplus sanguinipe
SwissProt_39:PR06_YEAST +  116.00  131.07  1.40      899  ! P19735 saccharomyces cerevis
SwissProt_39:DPOL_METJA +  116.00  126.19  1.44     1634  ! Q58295 methanococcus jannas
SwissProt_39:MAD1_YEAST +  115.50  131.96  1.49      749  ! P40957 saccharomyces cerevis
SwissProt_39:MNN4_YEAST +  115.50  128.27  1.53     1178  ! P36044 saccharomyces cerevi
SwissProt_39:NCPR_CANMA +  115.00  132.16  1.61      680  ! P50126 candida maltosa (yeas
SwissProt_39:POP1_YEAST +  115.00  130.10  1.62      875  ! P41812 saccharomyces cerevis
SwissProt_39:RPOB_PLAFA +  115.00  128.82  1.64     1024  ! P21421 plasmodium falciparu
SwissProt_39:RBP2_PLAVB +  115.00  127.19  1.65     1251  ! Q00799 plasmodium vivax (st
SwissProt_39:OPDA_HAEIN +  114.50  131.55  1.73      681  ! P44573 haemophilus influenza
SwissProt_39:YCF2_MARPO +  114.50  122.23  1.83     2136  ! P09975 marchantia polymorph
SwissProt_39:TOXB_CLODI +  114.50  121.40  1.83     2366  ! P18177 clostridium difficil
SwissProt_39:YCF1_EPIVI +  114.00  123.32  1.95     1738  ! Q00383 epifagus virginiana
SwissProt_39:YC82_EUGGR -  112.50  131.45  2.32      516  ! P30397 euglena gracilis. hyp
SwissProt_39:TOPG_SULAC +  112.50  124.24  2.41     1248  ! Q08582 sulfolobus acidocald
SwissProt_39:ATX1_PLAFA +  112.50  120.58  2.46     1956  ! Q04956 plasmodium falciparu
SwissProt_39:T3MO_ECOLI +  112.00  129.03  2.53      645  ! P12364 escherichia coli. typ
SwissProt_39:CHI1_BACCI +  111.00  127.19  2.95      699  ! P20533 bacillus circulans. c
SwissProt_39:KCC4_YEAST +  111.00  125.00  2.99      915  ! P25389 saccharomyces cerevis
SwissProt_39:DP3A_UREPA +  110.00  123.34  3.49      969  ! Q9pq74 ureaplasma parvum (ur
SwissProt_39:Y228_BORBU +  110.00  123.33  3.49      971  ! O51246 borrelia burgdorferi
SwissProt_39:Y414_MYCGE +  109.50  122.20  3.78     1036  ! P47653 mycoplasma genitaliu
SwissProt_39:RI15_YEAST +  109.50  117.84  3.87     1770  ! P43565 saccharomyces cerevi
SwissProt_39:MTS1_STRSA +  109.00  125.38  3.99      653  ! P29347 streptococcus sang
SwissProt_39:XYLR_BACSU +  108.50  129.11  4.20      384  ! P16557 bacillus subtilis.
SwissProt_39:YP71_MYCMY +  108.50  128.52  4.21      413  ! P55802 mycoplasma mycoide
SwissProt_39:NU5M_MYXGL -  108.50  125.46  4.29      601  ! Q9g2w8 myxine glutinosa (
SwissProt_39:MTF1_FLAOK +  108.50  124.86  4.30      647  ! P14871 flavobacterium oke

seq_name: SwissProt_39:CHIT_BRUMA

seq_documentation_block:
ID   CHIT_BRUMA     STANDARD;      PRT;   504 AA.
AC   P29030;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MF1 ANTIGEN).
OS   Brugia malayi.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC   Onchocercidae; Brugia.
OX   NCBI_TaxID=6279;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92179220; PubMed=1542646;
RA   Fuhrman J.A., Lane W.S., Smith R.F., Piessens W.F., Perler F.B.;
RT   "Transmission-blocking antibodies recognize microfilarial chitinase
RT   in brugian lymphatic filariasis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:1548-1552(1992).
CC   -!- FUNCTION: THE MF1 ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY
CC       FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-
CC       FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND
CC       TRANSMISSION.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MF1 ANTIGEN CORRESPOND
CC       WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.
CC   -!- PTM: O-GLYCOSYLATED.
CC   -!- MISCELLANEOUS: KNOWN TO BIND CALCIUM.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73689; AAA27854.1; -.
DR   PIR; A38221; A38221.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   Pfam; PF01607; Chitin_bind_2; 1.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   SMART; SM00494; ChtBD2; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Signal; Antigen;
KW   Repeat; Glycoprotein; Calcium-binding.
FT   SIGNAL        1     22
FT   CHAIN        23    504       ENDOCHITINASE.
FT   DOMAIN       23    400       CATALYTIC.
FT   DOMAIN      401    450       SER/THR-RICH (LINKER).
FT   DOMAIN      407    448       3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT   ACT_SITE    148    148       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   504 AA;  55971 MW;  A78BE7BFB8E3709B CRC64;


alignment_scores:
            Quality:  145.00              Length:    445
              Ratio:  0.687                Gaps:     26
 Percent Similarity:  47.416   Percent Identity: 22.247

alignment_block:
```

```
US-09-579-383-2 x CHIT_BRUMA  ..

Align seg 1/1  to: CHIT_BRUMA  from: 1  to: 504

    313 GGTATTATAGCAGGA...TACTATGGTTCATGG............AACAG 347
        |||   :::   |||   ||||||   :::|||           :::::
     22 GlyTyrValArgGlyCysTyrTyrThrAsnTrpAlaGlnTyrArgAspGl 38

    348 TCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAA 397
        ::::|||                  |||   ::::::      :::   |
     38 yGluGlyLysPheLeuProGlyAsnIleProAsnGlyLeuCysThrHisI 55

    398 TTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGA 447
        ||||||||   |||||||||:::::::::   |||   |||   |||:::
     55 leLeuTyr...AlaPheAlaLysValAspGluLeuGlyAsp...SerLys 69

    448 CCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATA 497
        ||||||:::   :::::::::                          :::
     70 ProPheGluTrpAsnAspGlu.....................AspTh 78

    498 TGAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTAC 547
         |||            |||   :::   :::   ::::::|||::::::
     78 rGluTrpSerLysGlyMetTyrSerAlaValThrLysLeuArgGluThrA 95

    548 GTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTAT...... 591
        ::|||   :::   :::|||||||||   ||||||   ::::::
     95 snProGlyLeuLysValLeuLeuSerTyrGlyGlyTyrAsnPheGlySer 111

    592 ......ATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAATATT 635
              :::      |||   |||:::   :::   ::::::::::::::
    112 AlaIlePheThrGlyIleAlaLysSerAlaGlnLysThrGluArgPheIl 128

    636 GAAGCTTGTTAATGATTTT............GATTTAGATGGTGTAGATA 673
        :|||   :::      |||            ::::::||||||   |||:
    128 eLysSerAlaIleAlaPheLeuArgLysAsnAsnPheAspGlyPheAspL 145

    674 TTGACTGGGAA...CCACATGGGAAGTTTTACAACTTAAATGAATTAAAT 720
        ::||||||||||   |||   |||
    145 euAspTrpGluTyrProValGly........................... 152

    721 TTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCC 770
          ::::::   :::   ||||||::::::   ::::::   ::::::
    153 ValAlaGluGluHisAlaLysLeuValGluAlaMetLysThrAlaPheVa 169

    771 GGAAGAAAAGTTAATTTCAATTTCTGGTTCATCA............AATG 808
         ||||||          :::   ||||||:::::            :::|
    169 lGluGlu......AlaLysThrSerGlyLysGlnArgLeuLeuLeuThrA 184

    809 CTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAA 858
        |||||:::|||:::   :::|||:::   :::          :::   |||
    184 laAlaValSerAlaGlyLysGlyThrIleAspGlySerTyrAsnValGlu 200

    859 TCT.........CCATATAACACTAAATTTTTG.................. 882
        |||            ::::::      ||||||
    201 SerLeuGlyLysAsnPheAspLeuLeuPheLeuMetSerTyrAspLeuHi 217

    883 ..........TCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAG 922
                 ::::::::::::::   :::   :::||||||   :::
    217 sGlySerTrpGluLysAsnValAspLeuHisGlyLysLeuHisProThrL 234

    923 CAGCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATACA...... 966
          :::   :::|||:::         |||:::|||||||||||||
    234 ysGlyGluValSerGly.........IleGlyIlePheAsnThrGluPhe 247

    967 ............................GCAAAGGAGAAAATAGATCT 986
                                    ||||||||||||   ::
    248 AlaAlaAspTyrTrpAlaSerLysGlyMetProLysGluLysIleIleIl 264

    987 T.........GTATTTATTCAAACATACAATTTAGAAACT.......... 1017
        :         ::::::   |||   ::::::|||::::::
    264 eGlyIleProMetTyrAlaGlnGlyTrpThrLeuAspAsnProSerGluT 281

   1018 .........................................ACAAAT 1023

                                                    ||||||
    281 hrAlaIleGlyAlaAlaAlaSerArgProSerSerAlaSerLysThrAsn 297

   1024 CCAGATATAATGGTAGATATGTAC.........TTATCCCATTTATATTT 1064
        |||          :::      |||             ::::::|||
    298 ProAlaGlyGlyThrAlaSerTyrTrpGluIleCysLysTyrLeuLysGl 314

   1065 TGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAGAACATA 1114
         |||   |||   ::::::                          |||:
    314 uGlyGlyLysGluThrVal.........................HisG 322

   1115 ACAGAGGT............................GGATTTAGTCCC 1134
        ::::::|||                                ::::::
    322 lnGluGlyValGlyAlaTyrMetValLysGlyAspGlnTrpTyrGlyTyr 338

   1135 GAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAA 1184
        :::|||:::|||   :::::::::      |||   :::   :::|||::
    339 AspAsnGluGluThrIleArgIleLysMetLysTrpLeuLysGluLysGl 355

   1185 TCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTA....... 1227
        :   :::                |||      ||||||   |||
    355 yTyrGly..............GlyAlaPheIleTrpAlaLeuAspPheA 367

   1228 .....TTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATATT 1272
              |||      ::::::         |||   :::   :::
    367 spAspPheThrGlyLysSerCysGlyLysGlyProTyrProLeu...... 381

   1273 TTTCTTACAAATATTTGGAAACATTTA..............AATCCTGA 1307
           |||:::   |||   ::::::|||              ||||||||
    382 ...LeuAsnAlaIleSerSerGluLeuGluGlyGluSerGluAsnProGl 397

   1308 AGTACAAACTCCAAAA...GACCTTACTATAACTGAA.............. 1341
        |:::   |||   :::   :::::::|||   ||||||
    397 uIleThrThrGluGluProSerIleThrGluThrGluAlaTyrGluThrA 414

   1342 ..AACCCTGAAGACTGTAGCACAATAGATGAATAT 1374
          :::   |||:::   |||      :::   |||
    414 spGluThrGluGluThrSerGluThrGluAlaTyr 425
```

```
seq_name: SwissProt_39:CHIA_SERMA

seq_documentation_block:
ID   CHIA_SERMA     STANDARD;      PRT;   563 AA.
AC   P07254; Q54275;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14).
GN   CHIA.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
OX   NCBI_TaxID=615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Koo J.C., Lim C.O., Choi Y.J., Kim C.Y., Bahk J.D., Lee S.Y.,
RA   Cho M.J.;
RL   Submitted (JAN-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RA   Jones J.D.G., Grady K.L., Suslow T.V., Bedbrook J.R.;
RT   "Isolation and characterization of genes encoding two chitinase
RT   enzymes from Serratia marcescens.";
RL   EMBO J. 5:467-473(1986).
RN   [3]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 24-31.
RC   STRAIN=BJL200;
RX   MEDLINE=95154677; PubMed=7851747;
RA   Brurberg M.B., Eijsink V.G.H., Nes I.F.;
RT   "Characterization of a chitinase gene (chiA) from Serratia marcescens
RT   BJL200 and one-step purification of the gene product.";
RL   FEMS Microbiol. Lett. 124:399-404(1994).
```

```
RN   [4]
RP   REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=95219379; PubMed=7704527;
RA   Perrakis A., Tews I., Dauter Z., Oppenheim A.B., Chet I., Wilson K.S.,
RA   Vorgias C.E.;
RT   "Crystal structure of a bacterial chitinase at 2.3-A resolution.";
RL   Structure 2:1169-1180(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L01455; AAA26551.1; -.
DR   EMBL; X03657; CAA27292.1; -.
DR   EMBL; Z36294; CAA85291.1; -.
DR   PIR; A25090; A25090.
DR   PDB; 1CTN; 20-JUL-95.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Signal; 3D-structure.
FT   SIGNAL       1     23
FT   CHAIN       24    563       CHITINASE A.
FT   DOMAIN      150    563       CATALYTIC.
FT   ACT_SITE    315    315       PROTON DONOR (PROBABLE).
FT   ACT_SITE    391    391       PROBABLE.
FT   CONFLICT     52     52       N -> S (IN REF. 3).
FT   CONFLICT     73     73       A -> T (IN REF. 3).
FT   CONFLICT     76     77       TA -> GP (IN REF. 2).
FT   CONFLICT     79     79       I -> V (IN REF. 3).
FT   CONFLICT    121    121       T -> S (IN REF. 3).
FT   CONFLICT    139    139       A -> P (IN REF. 2).
FT   CONFLICT    226    226       V -> I (IN REF. 2).
FT   CONFLICT    395    395       P -> A (IN REF. 1, 2 AND 3).
FT   CONFLICT    410    430       PAWKPDTAYTTVNGVNALLAQ -> RPGSRHRLHHGERRQC
FT                                AAGQ (IN REF. 1 AND 2).
FT   CONFLICT    437    437       V -> I (IN REF. 2 AND 3).
FT   CONFLICT    464    467       ATGP -> HRA (IN REF. 2).
FT   CONFLICT    473    473       K -> E (IN REF. 2 AND 3).
FT   CONFLICT    484    484       G -> S (IN REF. 2).
SQ   SEQUENCE   563 AA;  60979 MW;  0696FEF6AF83AA35 CRC64;


alignment_scores:
           Quality:  143.50            Length:    425
             Ratio:  0.721              Gaps:     20
 Percent Similarity:  46.824   Percent Identity: 20.471

alignment_block:
 US-09-579-383-2 x CHIA_SERMA   ..

 Align seg 1/1  to: CHIA_SERMA  from: 1  to: 563

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
```

```
         :::::      |||             ::::::::::::      :::
  44 spGlnAlaAlaThr.............AlaTyrAsnAsnLeuValLysVal 56

 166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
     |||:::      :::      |||:::::::::   :::   ||||||
  57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

 214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA................ 246
      |||   :::::::::|||      |||:::|||
  73 aGlyThrThrAlaLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

 247 ...............................AATATAAATAAAAATGAT 264
                                     ::::::||||||:::
  90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

 265 AGA.......................................... 267
     |||
 107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

 267 ........................................... 267

 123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuAlaP 140

 268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
          |||:::|||    :::||||||:::   |||      ||||||
 140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

 313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
     :::         ::::::::::::|||:::   :::|||:::      ||
 155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

 354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
     |         :::      :::|||:::   |||      ::::::   :
 171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

 404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
     :::::   ::::::   |||               |||   :::   |||
 186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

 439 ................................GTATCTAGACCATTTAA 455
                                     ::::::|||      ::
 203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

 456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
     :||||||:::   |||   :::      |||           :::|||:
 219 nGlyArgGluAspPheLysValSerIleHisAspProPheAlaAlaLeuG 236

 494 AATATGAAACCTATGGTATGATG........................... 516
     ::           |||:::
 236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

 517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
     ::::::::::::      ::::::::::::::::||||||:::   |||||
 253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

 567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
     |   |||:::||||||   |||                 :::|||   |
 269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

 599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
     ||      |||:::      :::|||   ::::::   ::::::::::::
 286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

 649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
        :::      :::|||||||||||||||||||||||||   |||   ||
 303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

 693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
     ||||            ||||||::::::   :::   ::::::   |||:
 319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

 737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
     ::   |||::::::   ||||||      :::   :::
```

```
     336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln..... 348

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                 ||||||:::   :::|||
     349 .........................LeuSerAlaGluThrGly..... 354

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                        |||:::   |||:
     355 ................................ArgLysTyrGluLeuT 360

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         :::::   |||::::::::::|||:::           :::|||
     360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                ::::|||:::|||:::::::::::||
     375 .......................TyrAsnValAlaGlnAsnSerMetAs 383

     984 TCTTGTATTTATTCAAACATACAAT 1008
         |   :::|||::::::::::|||:::
     383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: SwissProt_39:NU5M_TRYBB

seq_documentation_block:
ID   NU5M_TRYBB     STANDARD;      PRT;   590 AA.
AC   P04540;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN   ND5.
OS   Trypanosoma brucei brucei.
OG   Mitochondrion.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85037915; PubMed=6093040;
RA   Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,
RA   van Boom J.H., Benne R.;
RT   "The sequence of the gene for cytochrome c oxidase subunit I, a
RT   frameshift containing gene for cytochrome c oxidase subunit II and
RT   seven unassigned reading frames in Trypanosoma brucei mitochrondrial
RT   maxi-circle DNA.";
RL   Nucleic Acids Res. 12:7327-7344(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=164;
RX   MEDLINE=87201680; PubMed=2437452;
RA   Jasmer D.P., Feagin J.E., Payne M., Stuart K.;
RT   "Variation of G-rich mitochondrial transcripts among stocks of
RT   Trypanosoma brucei.";
RL   Mol. Biochem. Parasitol. 22:259-272(1987).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X01094; CAB57807.1; -.
DR   EMBL; M14820; AAB59225.1; -.
DR   PIR; A04519; QQUTC5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   590 AA;  71494 MW;  CAE231AD9DFF5282 CRC64;
```

```
alignment_scores:
           Quality:  136.50            Length:    661
             Ratio:  0.533               Gaps:     35
 Percent Similarity:  38.729   Percent Identity: 19.062

alignment_block:
 US-09-579-383-2/rev x NU5M_TRYBB   ..

 Align seg 1/1  to: NU5M_TRYBB  from: 1  to: 590

    1696 TTTTATTCCATTCTTCTATCCCTGGACATTTGGATGCATCTAGTTTTGTG 1647
         |||   |||:::   |||||||||   :::   :::   :::||||||||
      25 PheLeuSerPheTrpLeuSerLeuValMetIleIlePheIleValLeuCy 41

    1646 TATG.....................ACTCTAGTGCCTG..... 1630
         ||||                     :::::: ||||||
      41 sMetIlePheSerPheLeuMetValSerVal.CysLeuTyrGlyTyrTyr 57

    1629 ......ACCTTCCGGAGGTCCTTGTTGCCACCATTTAATTAAATATGGTT 1586
                 |||         |||:::      |||      ||||||::
      58 TyrTyrAspPheCysLeuIleLeuMetLeuAspPheCysPheIleTrpLe 74

    1585 CCCCTTTCCATATAATAATAGAGCTTTCTTTATAGTCAGTGTTATA...A 1539
         :   :::      ::::::      :::||||||:::   :::|||   :
      74 uThrTyrValCysSerGlyPheTyrMetPheIleMetLeuLeuIleAsnM 91

    1538 TAATGGGCTGCTTTCCCATCGCATAT...TTTTTCATAGCATACTTTGAC 1492
         :::::   ||||||   :::   |||   |||:::      |||||||||
      91 etValPheCysPheIleValPheTyrAlaPheTyrTyrMetTyrPheAsp 107

    1491 CAA..............GTCCCATTCATATCTGTCTACACCAGGTGCAT 1457
         :::                :::   ::::::   :::::::::   |||||
     108 MetLeuLeuGlyArgPheLeuIleIlePheTrpIlePheValValCysMe 124

    1456 GAATTGAATAAGATCTAGTTTTCCATAT..................... 1429
         ||||   :::      |||:::   :::
     124 tAsnLeuPheIleLeuSerTyrAspPheLeuThrAlaTyrCysGlyTrpG 141

    1428 ........AGCATCATTGTGTTTGTAATATATCCCTATGGTTGGAATAAC 1387
                 :::::::::   |||   :::      |||   |||
     141 luLeuLeuGlyLeuPheSerPhePheLeuIleSerTyrPheTrp..... 155

    1386 GAGTCCTGGAACATATTCATCTAT........................ 1363
                ::::::::::|||
     156 ......TyrArgPheAlaLeuLysPheGlyPheLysAlaPhePheIl 170

    1362 .......TGTGCTACAGTCTTCAGGGTTTTCAGTTATAGTAAGGTCTTTT 1320
                  :::   |||:::   :::|||:::::::|||   :::|||
     170 eGlyLysIleGlyAspValLeuLeuIlePheAlaPheSerIleIlePheL 187

    1319 GGAGTTTGTACTTCAGGATTTAAATGTTTCCAAATATT.........TGT 1279
           |||      |||   :::      |||      :::          |||
     187 euSerAsnGlyPheCysMetThrThrPheTyrPheLeuAsnPhePheCys 203

    1278 AAGAAAAATATCTACATCGAATGATCCAGTTGGTAATTGTTC........ 1237
                    |||||||||   |||         |||
     204 MetAspTyrTyrTyrIleGluPheSerIleCysLeuLeuValGlyCysAl 220

    1236 .TTTCATAAATAAATG............................... 1222
          |||   |||
     220 aPheThrLysSerThrGlnPheGlyLeuHisIleTrpLeuProAspAlaM 237

    1221 ........CCATATCCCTATACC......................... 1207
                 |||   |||   :::
     237 etGluGlyProIleProValSerAlaLeuIleHisAlaAlaThrLeuVal 253

    1206 ATCTGCCCTATTATTATTTTGATTTTTATCATGTATTGTTTTTCCTACCA 1157
         :::|||   ||||||:::|||   |||:::   :::||||||:::::
     254 ValCysGlyIleIleLeuLeuSerPheValTyrTrpCysPheAspPheTr 270

    1156 ATTCTAATAATTCTTT...ATTTTCGGGACTAAATCC............. 1123
          |||      |||:::   ::::::|||   ::::::
     270 pPheSerTyrPheTyrAsnLeuIleGlyTrpSerThrLeuIleLeuIleL 287
```

```
1122 .....ACCTCTGTTATGTTCTAATGAAAAACCTAATATGATTGTGATGTT 1078
         |||      :::|||               |||:::   ||||||
 287 euMetThrLeuCysValPhe.............TyrAsnPheAspVal 298

1077 ATATTTTAAACCAAAATATAAATGGGATAAGTACATATCTACCATTATAT 1028
                                                :::|||::
 299 ....................................LysArgTyrVa 302

1027 CTGGATTTGTAGTTTCTAAATTGTATGTTTGAATAAATACAAGATCTATT 978
     :         :::|||      :::|||      ::::::::::::::
 302 l.........AlaPheSerThrIleCysGlnIleSerPheSerMetPheC 316

 977 TTCTCCTTTGCTGTATTAAAAATATTAATAAAAGTTCCTGCTGATAACAT 928
         ||||||           |||:::::::
 316 ysCysLeuCysIleAspIleTyrIleGlySer.................. 326

 927 CGCTGCTGCCCTATGTAATTCTTTATTTGTTTCTATTTGTTCAGACAAAA 878
                     :::|||:::|||:::::::::|||   :::
 327 ................LeuPhePheCysTyrHisMetPheTyrLysAl 337

 877 A......TTTAGTGTT......ATATGG..................... 862
           |||   |||      ||||||
 337 aThrLeuPheIleValLeuGlyIleTrpIleHisIlePhePheGlyLeuG 354

 861 ..............AGATTCTTCATCTTTACAGAAAGATGCAACTCCTGA 826
                     ||||||::::::         |||  :::|||
 354 lnAspLeuArgCysTyrPhePheMetTyrPheCysGlyCys..ValLeuA 370

 825 AACGCATGATAATGCAGCATTTGATGAACCAGAAATTGAAATTAACTTTT 776
       |||:::::::::         :::   ::::::   ::::::   |||
 370 laArgLeuLeuLeuIlePheAlaIleLeuAsnSerCysSerIleTrpPhe 386

 775 CTTCCGGAATAGTTTTTCTTAACAAGTTAATTAATTTAATATAATAATTT 726
     |||
 387 Leu..................................................... 387

 725 GAAAAATTTAATTCATTTAAGTTGTAAAACTTCCCATGTGGTTCCCAGTC 676
                         |||
 388 .....................CysGly..................... 389

 675 AATATCTACACCATCTAAATCAAAATCATTAACAAGCTTCAATATTTTAT 626
                                                 ||||||
 390 ..............................................PheTyr 391

 625 CCACATAATCAATTTCTTTTTCTATATCTATCATATAGGTTTCTCCACCT 576
           :::::::::|||   |||   :::          |||   :::
 392 CysLysAspMetLeuLeuAlaLeuLeuMetLeuLeuSerPheTyrAsn.. 407

 575 AAGGATAAAAGAATAATTACATCTGGACGTACTTTTCTGATACGTCTAAT 526
       |||   |||   |||   ::::::::::::|||   :::
 408 ....IleIleGluPheLeuPheIleSerIleIlePheIlePhe....... 420

 525 TTCATTAAGCATCATACCATAGGTTTCATATTCTAAGCCGTGTTTTCTTA 476
                       |||   ::::::   :::   ||||||
 421 ......................PheThrMetIleTyrAsnTyrPheLeu 429

 475 ATAGGAATCTTTGTCTTCCATTAAATGGTCTAGATACATCATATAACATA 426
     :::   :::::::::|||                          :::::
 430 LeuPhePheLeuMetPhe.........................ValPh 437

 425 TTAATGCGAGCAAATGCAATATATAAAATTGACACCATTGGGTTTGAATC 376
     :   |||                              :::   |||   :
 437 eLysCys............................PheCysLeuValA 444

 375 AATCATATGTTTTGCTCTATCACCTTGACTGTTCCATGAACCATAGTATC 326
     ::         ||||||:::      :::|||   :::   :::   :::
 444 spCysLeuPheLeuLeuPheAspTyrGluCysCysLeuValTyrCysLeu 460

 325 CTGCTATAATAC......CTTGTTTCCTTTTTTTATACTCCTCTAAAATT 282
     :::      |||      ::::::|||:::|||:::      :::   ||
 461 IleSerLeuTyrMetCysIleLeuSerIlePhePheIle...IleAspPh 476

 281 TGTCTTGGTG.....................ATTTTCTATCATTTTTATT 253
     ||||   :::                     :::|||:::   |||:::|
 476 eValCysIlePheValPheSerSerTyrCysValPheTrpSerPhePheL 493

 252 TATATTTTTATTATTTTTTGTAATTCTAGAGTTACACCCATCTCCGCAAT 203
     ||   ||||||   |||
 493 euAsnPheTyrAsnPhe................................ 498

 202 AAGATCCATATTCTACAAAATTACTATTATTCGATTTAAGATGTGAAAAA 153

 498 ..................................................... 498

 152 GACTCATGTATTTCCGTTTGATGAGTTTTATTTTTATTTTCCCGTATTAT 103
                     ||||||         |||         :::::::
 499 ...............PheAspIleAlaIlePheValValPheLeuIleL 510

 102 TCCCAATGAATTATTTATAT.................TATTTTTTC... 74
     ::   :::::::::||||||                  |||||||||
 510 euSerValGlyPheLeuTyrTyrGlyCysLeuPhePheTyrPhePheAsn 526

  73 .........CTTTCAAGGTTCTGGAATTTG 53
             :::      ||||||::::::
 527 IleAspCysIleMetLeuPheTrpArgIle 536
```

seq_name: SwissProt_39:KTXA_KLULA

seq_documentation_block:

ID KTXA_KLULA STANDARD; PRT; 1146 AA.
AC P09805;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE KILLER TOXIN ALPHA/BETA SUBUNITS PRECURSOR (RF2 PROTEIN) [CONTAINS:
DE KILLER TOXIN ALPHA SUBUNIT; KILLER TOXIN BETA SUBUNIT (ENDOCHITINASE)
DE (EC 3.2.1.14)].
OS Kluyveromyces lactis (Yeast).
OG Plasmid pGKl-1.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 2359/152;
RA Sor F., Fukuhara H.;
RT "Structure of a linear plasmid of the yeast Kluyveromyces lactis;
RT compact organization of the killer genome.";
RL Curr. Genet. 9:147-155(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84297209; PubMed=6473099;
RA Stark M.J.R., Mileham A.J., Romanos M.A., Boyd A.;
RT "Nucleotide sequence and transcription analysis of a linear DNA
RT plasmid associated with the killer character of the yeast
RT Kluyveromyces lactis.";
RL Nucleic Acids Res. 12:6011-6030(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85037931; PubMed=6387625;
RA Tokunaga M., Kawamura A., Hishinuma F.;
RT "Cloning and nucleotide sequences of the linear DNA killer plasmids
RT from yeast.";
RL Nucleic Acids Res. 12:7581-7597(1984).
RN [4]
RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX MEDLINE=87004569; PubMed=3758030;
RA Stark M.J.R., Boyd A.;
RT "The killer toxin of Kluyveromyces lactis: characterization of the
RT toxin subunits and identification of the genes which encode them.";
RL EMBO J. 5:1995-2002(1986).
RN [5]
RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX MEDLINE=90259069; PubMed=2342564;
RA Bradshaw H.D. Jr.;

```
RT   "Killer toxins.";
RL   Nature 345:299-299(1990).
RN   [6]
RP   CHITINASE ACTIVITY OF ALPHA-SUBUNIT.
RX   MEDLINE=91301161; PubMed=2070799;
RA   Butler A.R., O'Donnell R.W., Martin V.J., Gooday G.W., Stark M.J.R.;
RT   "Kluyveromyces lactis toxin has an essential chitinase activity.";
RL   Eur. J. Biochem. 199:483-488(1991).
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH
CC       THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE
CC       TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE
CC       TOXIN) TO GAIN ENTRY INTO THE CELL.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,
CC       BETA AND GAMMA.
CC   -!- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-
CC       SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.
CC   -!- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07127; CAA30137.1; -.
DR   EMBL; X00762; CAA25334.1; -.
DR   EMBL; X01095; CAA25569.1; -.
DR   PIR; S07915; S07915.
DR   HSSP; P02877; 1HEV.
DR   InterPro; IPR001002; Chitin_bind.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR002482; LysM.
DR   Pfam; PF00187; chitin_binding; 1.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   PRINTS; PR00451; CHITINBINDNG.
DR   ProDom; PD000609; Chitin_bind; 1.
DR   SMART; SM00270; ChtBD1; 1.
DR   SMART; SM00257; LysM; 1.
DR   PROSITE; PS00026; CHITIN_BINDING; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Plasmid; Toxin; Glycoprotein; Signal; Hydrolase; Glycosidase;
KW   Chitin degradation.
FT   SIGNAL       1     17       POTENTIAL.
FT   PROPEP      18     29       POTENTIAL.
FT   CHAIN       30    892       KILLER TOXIN ALPHA SUBUNIT (PROBABLE).
FT   CHAIN      895   1146       KILLER TOXIN BETA SUBUNIT (PROBABLE).
FT   DOMAIN     310    363       CHITIN-BINDING (BY SIMILARITY).
FT   DOMAIN     364    635       CATALYTIC.
FT   ACT_SITE   495    495       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD   771    771       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   858    858       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   868    868       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   876    876       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   995    995       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1082   1082       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1117   1117       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE  1146 AA;  128936 MW;  BF4B1764EB465DE6 CRC64;


alignment_scores:
            Quality:  136.00              Length:    652
              Ratio:  0.500                Gaps:     31
 Percent Similarity:  41.718   Percent Identity: 19.325

alignment_block:
 US-09-579-383-2 x KTXA_KLULA   ..
```

```
Align seg 1/1  to: KTXA_KLULA  from: 1  to: 1146

    109 CGGGAAAATAAAAAT..............AAAACTCATCAAACGGAAAT 143
        :::   :::||||||              ||||||::::::::::::
    244 LysProSerLysAsnSerAspGlySerCysLysThrTyrLysValSerSe 260

    144 ACATGAGTCTTTTTCACAT.....................CTTAAATCGA 172
            ||||||   |||                        |||:::   |
    260 rGlyGluSerCysSerSerIleAlaValLysTyrTyrProLeuSerLeuA 277

    173 ATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGTAAC 222
        ||:::   :::      :::   |||:::|||         ||||||:::
    277 snAspIleGluAsnTyrAsnLysGlyAsnTyrGlyTrpLysGlyCysSer 293

    223 TCTAGAATTACAAAAAATAATAAAAATATAAATAAAAATGATAGAAAATC 272
        |||         :::::::::   |||:::      :::|||   :::::
    294 Ser......LeuGlnLysAspTyrAsnLeuCysValSerAspGlySerAl 308

    273 ACCAAGACAAATTTTA.............................. 288
        :||||||   :::
    308 aProArgProValSerAsnProIleAlaGluCysGlyProLeuAlaProG 325

    289 ..GAGGAGTATAAAAAAAGGAAACAA..................... 312
          |||:::|||:::   :::
    325 lyGluLysTyrAsnAlaLysCysProLeuAsnAlaCysCysSerGluPhe 341

    313 .........GGTATTATAGCAGGATAC..................... 330
                ||||:::         |||
    342 GlyPheCysGlyLeuThrLysAspTyrCysAspLysLysSerSerThrTh 358

    331 .....................................TATGGTTCAT 340
                                             |||||||||
    358 rGlyAlaProGlyThrAspGlyCysPheSerAsnCysGlyTyrGlySerT 375

    341 GG............AACAGTCAAGGTGATAGAGCAAAACATATGATTGAT 378
                     :::|||         :::      :::   :::|||
    375 hrSerAsnValLysSerSerThrPheLysLysIleAlaTyrTrpLeuAsp 391

    379 TCA................................AACCCAATGGT 392
        :::                                |||
    392 AlaLysAspLysLeuAlaMetAspProLysAsnIleProAsnGlyProTy 408

    393 GTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTAT 442
          :::||||||:::   ||||||::::::||||||     ::::::::::
    408 rAspIleLeuHisTyrAlaPheValAsnIleAsnSerAspPheSerIleA 425

    443 CTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTA 492
        ::            :::::::::   ||||||           :::
    425 spAspSerAlaPheSerLysSerAlaPheLeuLysValThrSerSerLys 441

    493 GAATATGAAACCTATGGTAT....GATGCTTAATGAAATTAGACGTATCA 538
        :::      :::::::|||      |||
    442 LysIleProSerPheGlyGlyTrpAspPhe..................... 451

    539 GAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACC 588
            |||||||||      :::|||::::::          :::||||||
    452 ....SerThrSerProSerThrTyrThrIle.........PheArgAsnA 464

    589 TATATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAA 638
                      ::::::||||||    ::::::::::::::::::::
    464 laValLysThrAspGlnAsnArgAsn.ThrPheAlaAsnAsnLeuIleAs 480

    639 GCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAA.... 684
        :::::::|||   ::::::||||||||||:::|||:::||||||||||
    480 nPheMetAsnLysTyrAsnLeuAspGlyIleAspLeuAspTrpGluTyrP 497

    685 ........CCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCA 726
                |||                        ::::::   ::: . :::
    497 roGlyAlaProAspIleProAspIleProAlaAspAspSerSerSerGly 513

    727 AATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGA 776
        :::   |||:::   ::::::::::||||||:::      :::|||:::
    514 SerAsnTyrLeuThrPheLeuLysLeuLeuLysGlyLysMetProSerGl 530
```

```
 777 AAAGTTAATTTCAATTTCTGGTTCATCA.................AATG 808
 530 yLysThrLeuSerIleAlaIleProSerSerTyrTrpTyrLeuLysAsnP 547

 809 CTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAA 858
 547 heProIleSerAspIleGlnAsnThrValAspTyr.............. 558

 859 TCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGA 908
 559 ............................MetValTyrMetThrTyrAs 565

 909 ATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTT 958
 565 pIleHisGlyIleTrpGluTyrGlyLysAlaAsnSerTyrIleAsnCysH 582

 959 TTAATACAGCAAAGGAGAAAATAGATCTTGTA.................. 990
 582 isThrProArgLysGluIleGluAspAlaIleLysMetLeuAspLysAla 598

 991 ....................................TTTATTCAAAC 1001
 599 GlyValLysPheAsnLysValPheGlyGlyValAlaAsnTyrGlyArgSe 615

1002 ATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTAT 1051
 615 rTyrLysMetValAsnThrAsn........................... 622

1052 CCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTT 1101
 623 ...................CysTyrAsnTyrGly......CysGlyPhe 630

1102 TCATTAGAACATAACAGAGGTGGATTTAGTCCCGAA.............. 1137
 631 GlnArgGlu.........GlyGlyAsnSerArgAspMetThrAsnThrPr 644

1138 .............AATAAAGAATTATTAGAATTGGTAGGAAAAACAATAC 1174
 644 oGlyValLeuSerAspSerGluIleIleAsp..............IleA 656

1175 ATGATAAAAATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCAT 1224
 656 spSerSerAspLysLysAsnAspArgTrpValAspThrAsnThrAspCys 672

1225 TTATTTATGAAA.................GAACAATTACCAACTGGATC 1256
 673 IlePheMetLysTyrAspGlyAsnSerValValSerTrpProLysSerAr 689

1257 ATTCGATGTAGAT.................................A 1270
 689 gTyrAspLeuGluAspMetPheLysAsnTyrGlyPheAlaGlyThrSerL 706

1271 TTTTTCTTACAAATATTTGGAAACAT.........TTAAATCCTGAAGTA 1311
 706 euTrpAlaAlaAsnTyrPheLysHisAspGluTrpLysAsnAspGluAsp 722

1312 CAAACTCCAAAAGACCTTACTATAACTGAAAACCCT.............. 1347
 723 AspAsnAsnAspAsp.........ThrGluAspProPheAspGluGluAs 736

1348 ................GAAGACTGTAGCACA..............ATAG 1366
 736 nValTyrPheAspValTyrAspCysLysAsnLysAlaGlyTyrAspLeuA 753

1367 ATGAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAA 1416
 753 spAsnProValTyrGlyCysArgLeuGluThr..............Ala 764

1417 CACAATGATGCTATATGGAAA...ACTAGATCT................. 1446
 765 IleAsnIleIleIleTrpAsnGlyThrGluSerValAsnThrValLeuAs 781
```

```
1447 .....................TATTCAATTCATGCACCTGGTGTAGACA 1474
 781 nIleLeuAsnAspTyrAspAsnTyrIleLysTyrTyrGluAlaLeuThrA 798

1475 GATATGAATGGGACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGG 1524
 798 rgAlaHisTyrAspSerValMetGluLysTyrGluLysTrpLeuPheGlu 814

1525 AAAGCAGCCCATTAT...TATAACACTGACTATAAAGAAAGCTCTATTAT 1571
 815 GluAspGlyTyrTyrThrTyrTyrThrAspValAspGlyAspAspIleIl 831

1572 TATA 1575
 831 eIle 832
```

```
seq_name: SwissProt_39:CYSP_PLAFA

seq_documentation_block:
ID   CYSP_PLAFA     STANDARD;      PRT;   569 AA.
AC   P25805;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   THROPHOZOITE CYSTEINE PROTEINASE PRECURSOR (EC 3.4.22.-) (TCP).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92228005; PubMed=1565129;
RA   Rosenthal P.J., Nelson R.G.;
RT   "Isolation and characterization of a cysteine proteinase gene of
RT   Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 51:143-152(1992).
CC   -!- FUNCTION: PROBABLY DEGRADES ERYTHROCYTE HEMOGLOBIN.
CC   -!- DEVELOPMENTAL STAGE: THROPHOZOITE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC       PAPAIN FAMILY OF THIOL PROTEASES.
CC   -!- SIMILARITY: STRONGEST SIMILARITY TO CATHEPSIN L.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81341; AAA29578.1; -.
DR   PIR; A45624; A45624.
DR   HSSP; P07858; 1CSB.
DR   MEROPS; C01.077; -.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Thiolprot_act_site.
DR   Pfam; PF00112; Peptidase_C1; 2.
DR   PRINTS; PR00705; PAPAIN.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW   Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?    332       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN       333    569       THROPHOZOITE CYSTEINE PROTEINASE.
FT   DOMAIN       64     70       POLY-ASN.
FT   ACT_SITE    357    357       BY SIMILARITY.
FT   ACT_SITE    488    488       BY SIMILARITY.
FT   ACT_SITE    533    533       BY SIMILARITY.
FT   DISULFID    354    395       BY SIMILARITY.
FT   CARBOHYD     58     58       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    127    127       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   569 AA;  66880 MW;  E5E762A3637F7C60 CRC64;


alignment_scores:
            Quality:  131.50              Length:    340
              Ratio:  0.743                Gaps:     14
 Percent Similarity:  52.059   Percent Identity: 21.176

alignment_block:
 US-09-579-383-2 x CYSP_PLAFA   ..

 Align seg 1/1  to: CYSP_PLAFA  from: 1  to: 569

       7 TTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTCTGCAAA 56
         |||   :::|||||||:::   :::|||:::   |||::::::
      36 PheValLeuSerIleTyrAlaPheIleThrPheIleIlePhe........ 49

      57 TTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAA 106
                                                 :::|||||||:
      50 .....................................CysIleGlyIleL 54

     107 TACGG...GAAAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCT 153
         ::        ||||||:::::::::|||
      54 euTyrPheThrAsnLysSerSerAlaHis..................... 63

     154 TTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTA 203
                       :::|||||||:::|||
      64 ...............AsnAsnAsnAsnAsn..................... 68

     204 TTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAA 253
                                      |||||||:::   :::::::
      69 ..............................LysAsnGluHisSerLeuL 75

     254 ATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAA 303
         ::|||::::::   :::      |||   :::|||||||:::|||||||||
      75 ysLysGluGluIleGluLeuArgValLeuLeuGluLysTyrLysLys 91

     304 AGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::|||:::||||||:::                 ::::::   :::::::::
      92 GlnLysAspGlyIleLeu............AsnGluSerSerAsnGluGl 104

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
          |||:::   |||:::   ::::::|||:::          :::
     104 uAspGluGluLysTyrThrLeuAsnSerGluThrTyrAsnAsnLysAsnA 121

     404 ATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTT 453
           ::::::      :::   :::   :::   :::   ::::::   :::
     121 snValSerAsnIleLysAsnAspSerIleLysSerLysLysGluGluTyr 137

     454 AATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAAC 503
            :::   :::|||:::||||||:::|||
     138 IleAsnLeuGluArgIleLeuLeuGluLys..................... 147

     504 CTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAG 553
          |||     ::::::||||||      :::   ::::::   |||   :
     148 .TyrLysLysPheIleAsnGlu......AsnAsnGluGluAsnArgLysG 162

     554 ATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATA 603
         :::::      :::|||            |||          :::   :::
     162 luLeuSerAsnIleLeuHisLysLeuLeuGluIleAsnLysLeuIleLeu 178

     604 GAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGAT.. 651
         :::::::|||   |||       ||||||   :::   |||:::||||||
     179 ArgGluGluLysAsp......AspLysLysValTyrLeuIleAsnAspAs 193

     652 .TTTGATTTAGATGGT...GTAGATATTGACTGGGAACCACATGGGAAGT 697
          :::|||      |||   :::::::|||      :::   :::   |||:
     193 nTyrAspGluLysGlyAlaLeuGluIleGlyMetAsnGluGluMetLysT 210

     698 TTTACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATT 747
         ::   :::   :::   :::|||      :::|||      |||::::::
     210 yrLysLysGluAspProIleAsnAsnIleLysTyrAlaSerLysPhePhe 226

     748 AACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGG 797
         ::::::::::::::        ::::::|||::: ::::::   :::
     227 LysPheMetLysGlu.........HisAsnLysVal.TyrLysAsnIleA 240

     798 TTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTA 847
           :::|||||      ::::::   |||:::      :::::::|||:::
     240 spGluGlnMetArgLysPheGluIlePheLysIleAsnTyrIleSerIle 256

     848 AAGATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAA 897
         |||       |||                          |||||||||
     257 LysAsnHisAsnLys.....................LeuAsnLysAsnAl 266

     898 ACAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTT 947
          :::   |||:::   ::::::|||         |||:::::::|||   |
     266 aMetTyrLysLysLysValAsnGlnPheSerAspTyrSerGluGluGluL 283

     948 TATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTC 997
         ||                ::::::      |||   :::|||::::::
     283 eu..............LysGluTyrPheLysThrLeuLeuHisValPro 294

     998 AAACATACAATTTAGAAA 1015
         :::|||   |||   |||
     295 AsnHisMetIleGluLys 300

seq_name: SwissProt_39:RPC1_PLAFA

seq_documentation_block:
ID   RPC1_PLAFA     STANDARD;      PRT;  2339 AA.
AC   P27625;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92018020; PubMed=1656254;
RA   Li W.B., Bzik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;
RT   "Characterization of the gene encoding the largest subunit of
RT   Plasmodium falciparum RNA polymerase III.";
RL   Mol. Biochem. Parasitol. 46:229-240(1991).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
CC       SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE
CC       III.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73770; AAA29729.1; -.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
DR   Pfam; PF00623; RNA_pol_A; 2.
DR   Pfam; PF01854; RNA_pol_A2; 1.
```

```
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW   Zinc-finger; Nuclear protein.
FT   ZN_FING      88    101       C3H-TYPE (POTENTIAL).
SQ   SEQUENCE   2339 AA;  272829 MW;  EDD899363086FD48 CRC64;


alignment_scores:
           Quality:  131.00             Length:    531
             Ratio:  0.524               Gaps:     27
 Percent Similarity:  47.081   Percent Identity: 19.962

alignment_block:
 US-09-579-383-2 x RPC1_PLAFA   ..

 Align seg 1/1  to: RPC1_PLAFA  from: 1  to: 2339

      55 AATTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAAT 104
         |||::::::::::      ::::::||||||   ||||||:::   :::
    1533 AsnAsnAsnAsnAsnAspAsnAsnAsnAsnAspAsnAsnAsnAsnAsn.. 1548

     105 AATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTT 154
             ::::::::::::|||:::::::::      ::::::|||:::::
    1549 ....AsnAsnSerAsnAsnAsnAsnTyrTyrTyrAsnLeuHisAspAspV 1564

     155 TTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTAT 204
           ::::::|||      :::|||   |||         |||   :::
    1564 alAsnAsnLeuGlyValThrAsnTyrAsnThrAsnIleTyrProAsnAsp 1580

     205 TGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAAA........ 246
         |||   :::|||      ::::::   |||:::||||||:::
    1581 Cys...AsnGlyIleTyrGluLysGluThrAsnAsnAsnGluLeuThrTh 1596

     246 ..................................................  246

    1596 rAsnSerAsnMetCysAspLysAsnAsnAspPheSerAspGluPhePheA 1613

     247 ..AATATAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAG 294
           |||||||||:::||||||       :::::::::|||:::   |||::
    1613 snAsnIleAsnGluAsnAspLeuLeu.TyrAspAsnLysTyrTyrArgGl 1629

     295 TATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAA 344
         :|||   |||:::                           |||   |||
    1629 nIlePheLysAsn.........................ValIleGlyP 1637

     345 CAGTCAAGGTGATAGAG.................CAAAACATATGATTG 376
           |||:::|||:::|||                 |||:::      |||
    1637 heValSerValPheGluTyrValGluSerTyrLysGlnHisTyrIleLeu 1653

     377 ATTCAA.........ACCCAATGGTGTCAATTTTATATATTGCATTTGCT 417
         :::               :::|||   :::|||   :::   :::|||
    1654 PheProTyrGluIleIleLysTrpThrSerPheLeuLeuGluTyrLeuTh 1670

     418 CGCATTAATATGTTATATGATGTATCTAGACCATTTAATG......GAAG 461
           :::|||         ::::::|||         :::       ::::
    1670 rGluIleIleProThrAsnIlePheLeuHisThrLysLeuSerLysLysG 1687

     462 ACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGTA 511
         ::|||   ::::::   :::||||||:::  :::   :::   |||
    1687 luLysProThrHisGlnLysAsnThrGly..LysMetLysIleTyr.... 1701

     512 TGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATT 561
           ::::::||||||::::::          ::::::      :::|||
    1702 .....IleGluGluIleLysLysTrpLeuPheIleLys......AlaIle 1714

     562 ATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGA 611
         :::   :::      :::             |||::::::   |||::
    1715 AsnIleTyrLysTyrPheSerPheLysLysSerIleGluLeuIleLysLy 1731

     612 AATTGATTATGTGGATAAAATATTGAAG..................... 639
         :   ||||||   :::   |||:::|||
    1731 sLysAspTyrPheAsnTyrIleIleLysAsnTyrAspIleSerHisArgT 1748

     640 ..CTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCA 687

         :::::::::|||:::
    1748 yrIleIleHisAspTyr................................ 1753

     688 CATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTATTATAT 737
               :::|||   ||||||::::::|||              |||   ::
    1754 ......SerPheIleAsnLeuLysGlnLeu............TyrLeuPh 1764

     738 TAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTT 787
         :   ::::::|||:::   :::                    |||   ||||
    1764 eIlePhePheAsnIleTyrLysTyrPhe............LysTyrIleS 1777

     788 CAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCA......... 828
         ||      |||::::::   :::::::::|||:::   |||
    1777 erThrProGlyAspAlaValGlySerIleSerAlaGlnSerIleGlyGlu 1793

     829 ..........................................GGAGTTGCATC 839
                                                 ||||||||||
    1794 ProGlyThrGlnMetThrLeuLysThrPheHisPheAlaGlyValAlaSe 1810

     840 TTTC..............................TGTAAAGATGAAG 856
         |:::                                 :::   :
    1810 rMetAsnValThrLeuGlyValProArgIleLysGluIleIleAsnAlaS 1827

     857 AATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAA 906
         :::::      :::|||   :::|||:::      :::|||:::|||
    1827 erAsnSerIleGlnThrProIleLeuAsnIleProLeuGluValAsnAsp 1843

     907 GAATTACATAGGGCAGCAGCGATG........................... 930
         :::   :::   |||      |||
    1844 AsnTyrAsnPheAlaLeuMetMetLysSerLysLeuGluLysThrThrIl 1860

     931 ................................TTATCAGCAGGAACTT 946
                                        |||   |||:::|
    1860 eArgAspIleCysMetTyrIleLysGluAspTyrThrSerArgGlyValP 1877

     947 TTATTAATATT...TTTAATACAGCAAAGGAGAAAATAGATCTTGTATTT 993
         ||:::::::::   ||||||      :::|||   |||:::   :::|||
    1877 heLeuSerValLysPheAsn......GluGluLeuIleGlnLysLeuPhe 1891

     994 ......ATTCAAACATACAATTTAGAAACTACAAATCCAGATATAATGGT 1037
               |||::::::|||||||:::::                |||:::
    1892 LeuAsnIleAsnAlaTyrAsnIleLysAspIle.........IleLeuLy 1905

    1038 AGATATGTACTTATCCCATTTATATTTTGGTTTAAAATATAACATCACAA 1087
          :::   :::::::::   :::   ::::::   |||   :::|||::::
    1905 sGlnSerHisIleAsnLysIleLysIleAsn...LysIleHisIleAsnV 1921

    1088 TCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAA 1137
         ::|||      :::::|||:::                ::|||   :::
    1921 alIleAsnLysTyrLysLeuHis..............IleSerLeuLys 1932

    1138 AATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCA 1187
         |||   |||::::::
    1933 AsnAspGluPheIle................................. 1937

    1188 AAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAAG 1237
                                                :::|||:::   |
    1938 ....................................PhePheGlnMetG 1942

    1238 AACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAAATATT 1287
         ||:::|||      |||   :::|||:::   |||:::
    1942 luSerLeuLysLysGlyLeuLeuAspLeuLeuIleTyrGlyAsp..... 1956

    1288 TGGAAACATTTAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAAC 1337
            |||   ::::::   :::      :::|||:::   :::||
    1957 ...LysAspIleLysArgCysIleIleLysLysGluAspIleGluValTh 1972

    1338 TGAAAACCCTGAAGAC...TGTAGCACAATAGATGAATAT 1374
         |:::|||   ::::::   |||:::   :::||||||||||
    1972 rAspAsnGluAspGluIleCysAspAspMetAspGluTyr 1985

seq_name: SwissProt_39:PTP1_DICDI
```

```
seq_documentation_block:
ID   PTP1_DICDI     STANDARD;      PRT;   521 AA.
AC   P34137;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROTEIN-TYROSINE PHOSPHATASE 1 (EC 3.1.3.48) (PROTEIN-TYROSINE-
DE   PHOSPHATE PHOSPHOHYDROLASE 1).
GN   PTPA OR PTP1.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93046662; PubMed=1423620;
RA   Howard P.K., Sefton B.M., Firtel R.A.;
RT   "Analysis of a spatially regulated phosphotyrosine phosphatase
RT   identifies tyrosine phosphorylation as a key regulatory pathway in
RT   Dictyostelium.";
RL   Cell 71:637-647(1992).
CC   -!- FUNCTION: MAY HAVE A ROLE IN GROWTH AND IN THE EARLY STAGES OF
CC       DEVELOPMENT. AFFECTS THE TIMING OF DEVELOPMENT.
CC   -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC       PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MIGHT BE TETHERED TO THE
CC       PLASMA MEMBRANE.
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN ANTERIOR-LIKE CELLS
CC       AND TO A LESSER DEGREE IN PRESTALK CELLS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT A VERY LOW LEVEL IN VEGETATIVE
CC       CELLS, INDUCED BY 4 HRS, MAXIMALLY EXPRESSED AT THE TIGHT
CC       AGGREGATE STAGE AND THROUGH THE REMAINDER OF DEVELOPMENT.
CC   -!- MISCELLANEOUS: THE PTPASE DOMAIN IS INTERRUPTED BY A PTPASE INSERT
CC       WHICH SHARES NO HOMOLOGIES WITH OTHER PTPASE PROTEINS.
CC   -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC       TYROSINE PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07125; AAA33241.1; -.
DR   HSSP; P28827; 1RPM.
DR   DictyDb; DD03018; ptpA.
DR   InterPro; IPR000340; DS_phosphatase.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_prot_phphtase.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00194; PTPc; 1.
DR   SMART; SM00012; PTPc_DSPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW   Hydrolase.
FT   DOMAIN        1    114       SER-RICH.
FT   ACT_SITE    310    310       BY SIMILARITY.
FT   DOMAIN      327    425       PTPASE INSERT (ASN-RICH).
FT   DOMAIN      382    400       POLY-ASN.
SQ   SEQUENCE   521 AA;  59427 MW;  0F516AEDD75EAB96 CRC64;


alignment_scores:
           Quality:  129.50            Length:    462
             Ratio:  0.583               Gaps:     23
 Percent Similarity:  48.052   Percent Identity: 20.563

alignment_block:
 US-09-579-383-2 x PTP1_DICDI   ..
```

```
Align seg 1/1  to: PTP1_DICDI  from: 1  to: 521

     64 ACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGA 113
        ::::::::::::::|||::::::|||:::|||:::
      3 SerValGluSerSerAsnGlnMetAsnGlySerIle.............. 14

    114 AAATAAAAATAAAACTCATCAAACGGAAATA..................... 144
            :::|||||||||:::::::   :::|||
     15 ....GluAsnLysThrAsnLysIleAspIleSerValLeuArgProLeuP 30

    145 ...........CATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAAT 183
                  :::::::|||:::|||       :::|||::::::|||:::
     30 roThrArgSerAsnSerSerIleSerLeuSerSerSerSerHisSerSer 46

    184 TTTGTAGAATATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTAC 233
        |||   :::   ||||||       :::          ||||||      ::
     47 PheSerArgMetGlySerLeuGlySerLeuProThrAsnSerGlySerSe 63

    234 AAAA.........AATAATAAAAATATAAAT..................... 255
        ::::        ||||||::::::::::::
     63 rSerProTyrTyrAsnAsnSerSerPheAspLeuValAspGluGluArgI 80

    256 ......................AAAAATGATAGAAAATCACCAAGACAA 282
                               ||||||       |||      ::::::
     80 leLysSerSerIleTyrAsnLeuLysAsnHisIleLysCysIleHisLys 96

    283 ATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTA 332
        |||   ||||||::::::      ::::::::::
     97 IleLysGluGluPheArgLeuLeuGluGluSerVal............Gl 109

    333 TGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAA 382
           |||   :::|||:::|||||:::   ||||||      ::::::|
    109 yProSerGluThrSerGluGlyAspLys...LysHisAsnThrSerLysA 125

    383 ACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTA 432
        ||      :::::||||||   :::      :::|||::::::::::
    125 snArgTyrThrAsnIleLeuProValAsnHisThrArgValGlnLeu... 140

    433 TATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAA 482
                                          :::  :::      ||
    141 ..................................LysLysIleGlnAspLy 146

    483 ACACGGCTTAGAATATGAA.............................. 501
        |:::|||   :::|||
    146 sGluGlySerAspTyrIleAsnAlaAsnTyrIleAspGlyAlaTyrProL 163

    502 ..............ACCTATGGTATGATGCTTAATGAAATT......... 528
                       |||   |||   :::   |||   |||
    163 ysGlnPheIleCysThrGlnGlyProLeuProAsnThrIleAlaAspPhe 179

    529 ...AGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTT 575
           |||   :::   :::   |||      ::::||::::::||||||
    180 TrpArgMetValTrpGluAsnArgCysArgIleIleValMetLeuSerAr 196

    576 AGGTGGAGAAACCTATATGATAGATATAGAAAAAGAAATTGATTATGTGG 625
           :::|||:::      |||      ::::::         |||
    196 gGluSerGluAsnCysArgIleLysCysAspArg.........TyrTrpP 210

    626 ATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATT 675
          :::                                       :::|||
    210 roGlu.....................................GlnIle 213

    676 GACTGGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTC 725
           |||      :::   :::   |||   ||||||:::   |||::
    214 GlyGlyGluGlnPheSerIleTyrGlyAsnGlyAsnGluVal...PheGl 229

    726 AAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAG 775
        :::|||   :::::::|||:::::::::::::::         ||||||:
    229 yThrTyrSerValGluLeuValGluValIleGlnAsp......ProGluA 244

    776 AAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTT 825
        :::::::::|||            :::   |||      |||:::
    244 rgGluIleIle............ThrArgAsnIleArgLeuThrPheGlu 256
```

```
     826 TCAGGAGTTGCATCTTTCTGTAAA.....................GATGA 854
         :::   :::   ::::::   :::                      |||::
     257 GlyGluThrArgAspIleThrGlnTyrGlnTyrGluGlyTrpProAspHi 273

     855 AGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAA......A 898
         ::::   |||   :::|||:::   :::   :::   ::::::      |
     273 sAsnIleProAspHisThrGlnProPheArgGlnLeuLeuHisSerIleT 290

     899 CAAATAAAGAA..................................... 909
         |||||::::::
     290 hrAsnArgGlnAsnGlnIleIleProSerSerAspArgAsnValProIle 306

     910 ...TTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATAT 956
            :::|||   :::|||::::::      :::||||||||||   :::
     307 IleValHisCysSerAlaGlyValGlyArgThrGlyThrPheCysThrAl 323

     957 TTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATACA 1006
                    |||   |||:::|||      |||   |||
     323 aValIleMetMetLys...LysLeuAspHisTyrPheLysGln....... 336

    1007 ATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTATCCCAT 1056
           |||:::::::|||      |||   :::|||||||   :::   :::|||
     337 ..LeuAspAlaThrProIleAspGlnValValAspProPhe...ThrHis 351

    1057 TTATATTTT.................................GGTTTAAA 1073
         |||   :::                                |||||
     352 LeuProIleThrGluTyrGlnSerAspAsnLeuAspLeuLysGlyLeuGl 368

    1074 ATATAACATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTG 1123
          |||::::::                |||      :::|||   :::|
     368 yTyrHisPheLys..............SerSerIleTyrAsnSerAsnG 380

    1124 GATTTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATA 1173
         ||::::::   :::|||                  ::::::::::::
     380 lyIleAsnAsnAsnAsn.....................AsnAsnAsnLeu 389

    1174 CATGATAAAAATCAAAATAATAATAGGGCAGATGGT 1209
         ::::::::::|||:::   ||||||::::::::::|||
     390 AsnAsnAsnAsnAsnIleAsnAsnAsnSerAsnGly 401
```

seq_name: SwissProt_39:CHIT_NPVAC

seq_documentation_block:

```
ID   CHIT_NPVAC     STANDARD;      PRT;   551 AA.
AC   P41684;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=46015;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C6;
RX   MEDLINE=94303173; PubMed=8030224;
RA   Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT   "The complete DNA sequence of Autographa californica nuclear
RT   polyhedrosis virus.";
RL   Virology 202:586-605(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L22858; AAA66756.1; -.
DR   HSSP; P07254; 1CTN.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR000886; ER_target.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS00014; ER_TARGET; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;
KW   Endoplasmic reticulum.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    551       PROBABLE ENDOCHITINASE.
FT   ACT_SITE    305   305       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    173   173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    444   444       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        548   551       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   551 AA;  61368 MW;  4DDAAD187873BBA2 CRC64;
```

alignment_scores:

```
            Quality:  129.50            Length:    417
              Ratio:  0.689               Gaps:     21
 Percent Similarity:  45.084   Percent Identity: 19.424
```

alignment_block:

US-09-579-383-2 x CHIT_NPVAC ..

Align seg 1/1 to: CHIT_NPVAC from: 1 to: 551

```
     286 TTAGAGGAGTATAAA......AAAAGGAAACAAGGTATTATAGCAGGATA 329
         |||||:::   |||      :::|||:::::   |||:::|||:::||
     137 LeuGluAsnAsnLysProGlyArgArgGluAspLysIleValAlaAlaTy 153

     330 CTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACATATGATTGATT 379
         |:::   :::|||:::      |||         :::      :::|||:
     153 rPheValGluTrpGlyValTyrGly......ArgAsnPheProValAspL 168

     380 CAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATT...... 423
         ::   |||:::  ::::::  ::::::  ::::::   |||
     168 ysValProLeuProAsnLeuSerHisLeuLeuTyrGlyPheIleProIle 184

     424 .....................................AATATGTT 431
                                              :::   ::
     185 CysGlyGlyAspGlyIleAsnAspAlaLeuLysThrIleProGlySerPh 201

     432 ATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAA 481
         :   ::::::::|||      :::|||||:::   |||   :::
     201 eGluSerLeuGlnArgSerCysLysGlyArgGluAspPheLysValAlaI 218

     482 AACAC..............................GGCTTAGAA... 495
           |||                             |||::::::
     218 leHisAspProTrpAlaAlaValGlnLysProGlnLysGlyValSerAla 234

     496 ..TATGAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAA 542
             |||   |||         ::::::::::::         :::
     235 TrpAsnGluProTyrLysGlyAsnPheGlyGlnLeuMetAlaAlaLysLe 251

     543 AGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACC.... 588
         ::::::|||   :::   ||||||   |||:::||||||   |||
     251 uAlaAsnProHisLeuLysIleLeuProSerIleGlyGlyTrpThrLeuS 268

     589 ..............TATATGATAGATATAGAAAAAGAAATTGATTATGTG 624
                        :::|||   |||:::||||||:::      :::|||
     268 erAspProPheTyrPheMetHisAspValGluLysArgAsnValPheVal 284

     625 GATAAAATATTGAAGCTTGTTAATGATTTTGAT...TTAGATGGTGTAGA 671
         |||::::::   ::::::::::::   :::      :::||||||||||||
     285 AspSerValLysGluPheLeuGlnValTrpLysPhePheAspGlyValAs 301
```

```
 672 TATTGACTGGGAA...CCACATGGGAAGTTTTACAAC......TTAAATG 712
     ||||||||||||||   |||   ||||||      |||      |||:::: 
 301 pIleAspTrpGluPheProGlyGlyLysGlyAlaAsnProSerLeuGlyA 318

 713 AATTAAAT...TTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGA 759
     ::   :::      ::::::   ||||||   |||::::::   |||||| 
 318 spAlaAspGlyAspAlaLysThrTyrIleLeuLeuLeuGluGluLeuArg 334

 760 AAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGC 809
           :::   :::                                     
 335 AlaMetLeuAspAsp..................................... 339

 810 TGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAAT 859
         |||::::::   :::|||                             
 340 ....LeuGluAlaGlnThrGly.............................. 345

 860 CTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAA 909
                                                     :::  
 346 ..................................................Arg 346

 910 TTACATAGGGCAGCAGCGATGTTATCAGCAGGA............ACTTT 947
     :::::::::   ::::::   :::|||||||||            :::  
 347 ValTyrGluLeuThrSerAlaIleSerAlaGlyTyrAspLysIleAlaVa 363

 948 TATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTC 997
      :::|||   :::      |||:::::::::::::      :::|||:::: 
 363 lValAsn...TyrAlaGluAlaGlnLysSerLeuGlyLysIlePheLeuM 379

 998 AAACATACAATTTAGAA......ACTACAAATCCAGATATA......... 1032
     :::::|||:::::::::          :::|||   |||:::          
 379 etSerTyrAspPheLysGlyAlaTrpSerAsnThrAspLeuGlyTyrGln 395

1033 ..........................ATGGTAGATATGTACTTATCCCA 1055
                                      ::::::|||   :::||  
 396 ThrThrValTyrAlaProSerTrpAsnSerGluGluLeuTyrThrThrHi 412

1056 TTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCAT 1105
     |              |||   :::      :::|||      ::::      
 412 s.................TyrAlaValAspAlaLeuLeuLysGlnGlyV 423

1106 TAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTAGAA 1155
     :::::   |||:::                                     
 423 alAspProAsnLys................................Ile 428

1156 TTGGTAGGAAAAACAATACATGATAAA..................... 1182
     :::||||||   :::::::::   :::                      
 429 IleValGlyValAlaMetTyrGlyArgGlyTrpThrGlyValThrAsnTy 445

1183 ....AATCAAAAT............AATAATAGGGCAGATGGTATAGGGA 1216
         |||:::|||           ::::::::::::   |||   |||  
 445 rThrAsnAspAsnTyrPheSerGlyThrGlyAsnGlyProGlySerGlyT 462

1217 TATGGCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTA 1266
       |||:::                                     |||  
 462 hrTrpGlu.....................................AspGly 466

1267 GATATTTTTCTTACAAATATTTGGAAACATTTAAATCCTGAAGTACAAAC 1316
        :::         :::|||   |||   ||||||      |||   ||  
 467 ValValAspTyrArgGlnIleGlnLysAspLeuAsnAsnTyrValTyrTh 483

1317 T 1317
     |
 483 r 483
```

seq_name: SwissProt_39:Y832_METJA

seq_documentation_block:

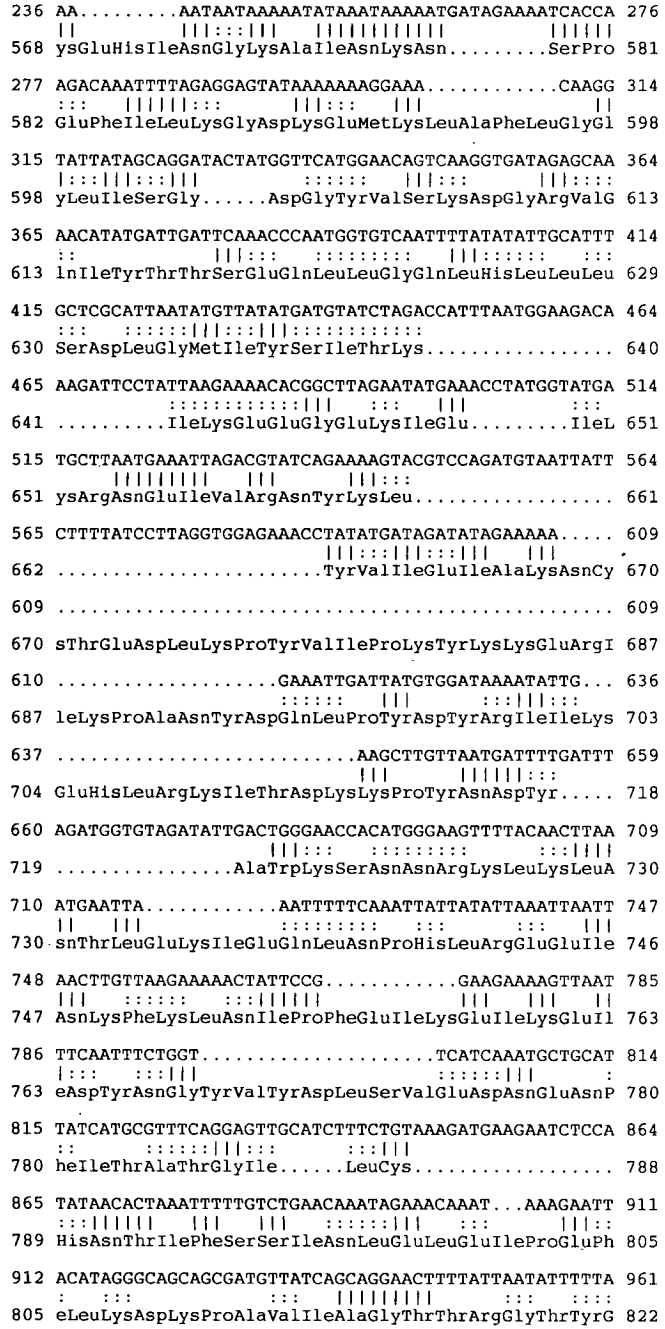
```
ID   Y832_METJA     STANDARD;      PRT;  1750 AA.
AC   Q58242;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0832.
GN   MJ0832.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC       A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC       FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC   -!- SIMILARITY: WEAK IN THE C-TERMINAL, TO M.JANNASCHII MJ0885.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67527; AAB98834.1; -.
DR   TIGR; MJ0832; -.
DR   InterPro; IPR003586; HintC.
DR   InterPro; IPR003587; HintN.
DR   InterPro; IPR002203; Intein.
DR   PRINTS; PR00379; INTEIN.
DR   SMART; SM00305; HintC; 2.
DR   SMART; SM00306; HintN; 2.
DR   PROSITE; PS00881; PROTEIN_SPLICING; FALSE_NEG.
KW   Hypothetical protein; Autocatalytic cleavage; Protein splicing;
KW   Complete proteome.
FT   CHAIN         1    337       HYPOTHETICAL PROTEIN MJ0832, 1ST PART
FT                                (POTENTIAL).
FT   CHAIN       338    790       MJA RNR-1 INTEIN (POTENTIAL).
FT   CHAIN       791   1058       HYPOTHETICAL PROTEIN MJ0832, 2ND PART
FT                                (POTENTIAL).
FT   CHAIN      1059   1591       MJA RNR-2 INTEIN (POTENTIAL).
FT   CHAIN      1592   1750       HYPOTHETICAL PROTEIN MJ0832, 3RD PART
FT                                (POTENTIAL).
SQ   SEQUENCE   1750 AA;  203279 MW;  6D6893B770A25359 CRC64;
```

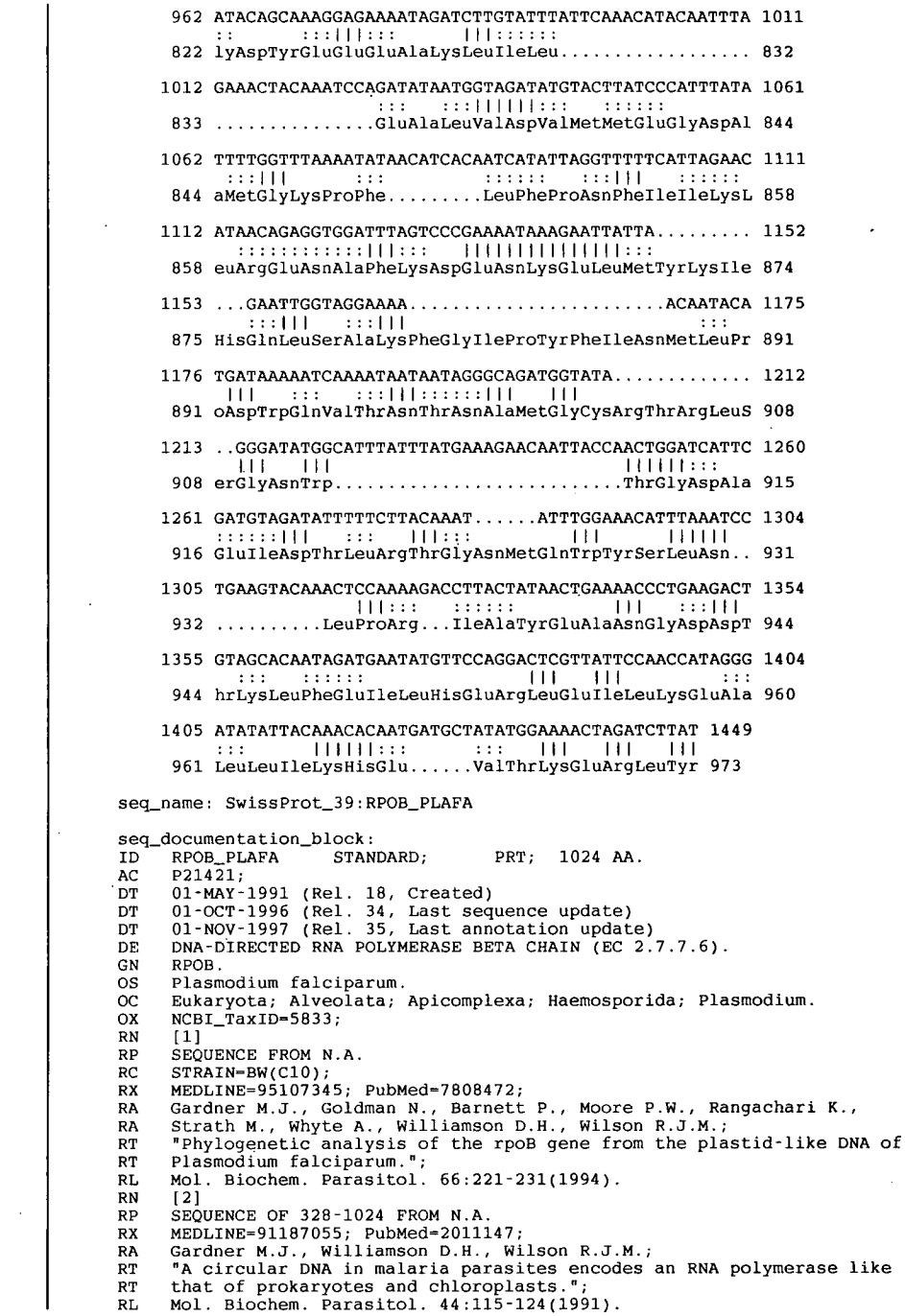
alignment_scores:

```
            Quality:  124.50            Length:    515
              Ratio:  0.530               Gaps:     28
 Percent Similarity:  45.631   Percent Identity: 20.777
```

alignment_block:

```
 US-09-579-383-2 x Y832_METJA   ..

 Align seg 1/1  to: Y832_METJA  from: 1  to: 1750

 139 GAAATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGT...AATTT 185
     :::|||:::|||:::::::::      :::   ::::::|||   :::|| 
 542 GlnIleAsnGluAsnIleAlaValLysArgTyrGluAspSerValArgPh 558

 186 TGTAGAATATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACAA 235
     ||||:::   |||   |||                     |||:::   |  
 558 eValAsnLysGlyPheTyr...................ArgPheLeuL 568
```





```
seq_name: SwissProt_39:RPOB_PLAFA

seq_documentation_block:
ID   RPOB_PLAFA     STANDARD;      PRT;  1024 AA.
AC   P21421;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6).
GN   RPOB.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BW(C10);
RX   MEDLINE=95107345; PubMed=7808472;
RA   Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RA   Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RT   "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
RT   Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 66:221-231(1994).
RN   [2]
RP   SEQUENCE OF 328-1024 FROM N.A.
RX   MEDLINE=91187055; PubMed=2011147;
RA   Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT   "A circular DNA in malaria parasites encodes an RNA polymerase like
RT   that of prokaryotes and chloroplasts.";
RL   Mol. Biochem. Parasitol. 44:115-124(1991).
```

```
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75544; CAA53232.1; -.
DR   EMBL; X52177; CAA36427.1; -.
DR   EMBL; X95275; CAA64572.1; -.
DR   PIR; S10438; RNZQBF.
DR   InterPro; IPR001572; RNA_pol_B.
DR   Pfam; PF00562; RNA_pol_B; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase.
SQ   SEQUENCE   1024 AA;  122185 MW;  B49FBDED7C9BD412 CRC64;


alignment_scores:
            Quality:  121.50              Length:    651
              Ratio:  0.406                Gaps:     37
 Percent Similarity:  45.929   Percent Identity: 20.737

alignment_block:
 US-09-579-383-2/rev x RPOB_PLAFA    ..

 Align seg 1/1  to: RPOB_PLAFA  from: 1  to: 1024

    1762 ATTGTAATGGTAAATCCACTTCTTGTT.........CATATTGTTCCTCT 1722
         |||   :::||||||||||:::||||||         :::|||:::
       2 IleTyrIleValAsnProIleLeuValLysAsnAsnTyrIleIleSerAs 18

    1721 ACTTCTAGTGGTTTATGTGGAT........................... 1700
          |||   ::::::   ::::::
      18 nLeuTyrLeuLeuLeuIleGlnGluIleIleTyrAsnLeuArgTyrTyrI 35

    1699 .....................................ATTTTTTAT 1691
                                              |||:::|||
      35 leLeuPheLeuAsnAsnAsnIleAsnValLysPheAsnPheIleTyrTyr 51

    1690 TCCATTCTTCTA.......................TCCCTGGACATTTG 1665
         :::|||::::::                       |||:::|||
      52 LysIleIleIleLeuLeuThrAsnIleAsnIleAsnSerIleAspThrIl 68

    1664 GATGCATCTAGTTTTGTGTATGACTCTAGTGCCTGACCTTCCGGAGGTCC 1615
          ::::::::::         :::   ::::::  ||||||::::::   :::
      68 eGlnAsnIleAsnAsnLeuLeuLysIleIle.LeuThrLeuLysLeuAsn 84

    1614 TTGTTGCCA..............CCATTTAATTAAATATGGTTC..... 1585
         ::::::                  ||||||   :::   |||
      85 PheIleAsnIleAsnLysIleIleLysPheAsnIleLeuIlePheIleLe 101

    1584 .CCCTTTCCATATAATAATAGAGCTTTCTTTATAGTCAGTGTTATAATAA 1536
          ||||||   |||||||||                 :::|||||||:
     101 uProPheIleTyrAsnAsn.....................IleIleIleL 111

    1535 TGGGCTGCTTTCCCATCGCATATTTTTTCATAGCATACTTTGACCAAGTC 1486
         :::::
     111 euAsn..................................................... 112

    1485 CCATTCATATCTGTCTACACCAGGTGCATGAATTGAATAAGATCTAGTTT 1436
                 :::::::|||      |||::::::   ::::::::::::
     113 .........GlyLeuTyrLysThrCysIleGlnLeuPheLysLysAsnAs 126

    1435 TCCATATAGCATCATTGTGTTT.................GTAATATATC 1404
                  ||||||   |||                    :::||||||
      126 nLysIlePheIleIleLysPheLysAsnAsnAsnLysAsnIleIleTyr. 142

    1403 CCTATGGTTGGAATAACGAGTCCTGGAACATATTCATCTATTGTGCTACA 1354
                                        ::::::||||||   :::
      143 ............................ValTyrIleTyrIleSerLeu 149

    1353 GTCTTCAGG...GTTTTCAGTTATAGTAAGGTCTTTTGGAGTTTGTACTT 1307
           :::|||   :::|||:::   ||||||:::      :::   |||||
      150 GlyLeuArgIleIlePheLysIleSerLysLeuAsnIleAspCysTyrPh 166

    1306 CAGGATTTAAATGTTTCCAAATATTTGT............AAGAAAAATA 1269
         |:::         |||      :::                    :::|||||
      166 eAsnAsnPheLysPheAsnPheLeuIleLeuLeuLeuTyrLeuAsnAsnI 183

    1268 TCTACATCGAATGATCCAGTTGGTAATTGTTCTTTCATAAATAAATGCCA 1219
         |||||||| |||      ::::::          :::   |||
      183 leTyrIle.AsnLysAsnIleSerLeuPheIleTyrAsnAsn........ 196

    1218 TATCCCTATACCATCTGCCCTATTATTATTTTGATTTTTATCATGTATTG 1169
                |||   :::   :::|||::::::   :::::::::   |||
      197 ....IleIleAsnLysLysIleLeuIleTyrAsnTyrIleLysPheIleT 212

    1168 TTTTTCCTACCAATTCTAATAATTCTTTATTTTCGGGACTAAATCCACCT 1119
                     |||:::   |||:::::::::|||   |||:::
      212 yrSerLysTyrAsnAsnIleAsnAsnIleIleSer...LeuLysLeuPhe 227

    1118 CTGTTATGTTCTAATGAAAAACCTAATATGATTGTGATGTTA........ 1077
         ::::::      |||:::      |||:::   :::   |||
      228 IleIleLysLeuAsnLysPheAsnAsnIleTyrIleAsnLeuLeuAsnIl 244

    1076 .TATTTTAAACCAAAATATAAATGGGATAAGTACATATCTACCATTATAT 1028
             |||:::   |||   :::::::::   |||   :::      |||:
      244 eLeuPheSerIleLysLeuAsnPheSerTyrTyrSerAspPheTyrIleA 261

    1027 CTGGATTTGTAGTTTCTAAATTGTATGTTTGAATAAATACAAGATCTATT 978
         ::::::::      :::|||:::|||      |||::::::      |||
      261 snAsnIleTyrAsnLysLysPheTyrSerIleIleAspAsnLeuLeuIle 277

    977 TTCTCCTTTGCTGTATTAAAAATA............TTAATAAAAGTTCC 940
           |||          |||||||||            |||:::::::::
      278 LysSerLysLysTyrLeuLysIlePheLysTyrGlnLeuLeuAsnIleAs 294

    939 TGCT.........GATAACATCGCTGCTGCCCTATGTAATTCTTTATTTG 899
                     :::||||||:::       |||   |||:::   ::::
      294 nArgAsnIleTyrAsnAsnIleThrLeuLeuLeuAsnAsnLysLysTyrI 311

    898 TTTCTATTTGTTCAGAC............AAAAATTTAGTGTTATATGGA 861
         :::::|||      :::            :::   ||||||   |||:::
      311 leAsnIleIleLeuGluAsnIleAsnIleAsnProLeuValGlnTyrSer 327

    860 GATTCTTCATCTTTACAGAAAGATGCAACTCCTGAAACGCATGATAATGC 811
         |||:::                                    :::|||
      328 AspGlnVal................................AsnAsnLe 333

    810 AGCATTTGATGAACCAGAAATTGAAATTAACTTTTCTTCCGGAATAGTTT 761
         :::    :::||||||||||:::|||:::::::::::
      333 uSer.GluIleAsnGlnLysPheLysIleAsnMetIleThrThr...... 347

    760 TTCTTAACAAGTTAATTAATTTAATATAATAATTTGAAAAATTTA..... 716
           |||:::|||   :::|||      |||:::|||::::::|||
      348 GlyLeuAsnSerLysPheIleLeuAsnAsnAspLeuArgGluLeuProAr 364

    715 ..............ATTCATTTAAGTTGTAAAACTTCCCATGTGGTTC 682
                       |||   |||      :::|||::::::   :::
      364 gAsnIleLeuGlyTyrIleSerLeu...IleAsnThrAsnGluGlyLeuT 380

    681 CCAGTCAATATCTACACCATCTAAATCAAAATCATTAACAAGCTTCAATA 632
                  :::::::::|||:::   |||
      380 hrCysGlyLeuValAsnTyrLeuThrThrAsn.................. 390

    631 TTTTATCCACATAATCAATTTCTTTTTCTATATCTATCATATAGGTTTCT 582
```

```
                          :::||||||   |||:::|||   |||::
    391 .....................IlePheLeuAsnLeuLysTyrLeuPheVa 400

    581 CCACCTAAGGATAAAAGAATAATTACATCTGGACGTACTTTTCTGATACG 532
        :                    ::::::         :::|||
    400 l.....................IleTyrTyrLysHisIlePhe....... 407

    531 TCTAATTTCATTAAGCATCATACCATAGG......TTTCATATTCTAAGC 488
                           |||:::|||      |||   :::|||
    408 ....................TyrAsnArgTyrAsnPheLysLeuLeuLeu 417

    487 CGTGTTTTCTTAATAGGAATCTTTGTCTTCCATTAAATGGTCTAGATACA 438
        :::::::|||  |||:::|||:::
    418 AsnIlePhe..AsnLysAsnPhe........................... 424

    437 TCATATAACATATTAATGCGAGCAAATGCAATATATAAAATTGACACCAT 388
           |||||||||   :::      |||   |||||      :::||
    425 ...TyrAsnIleSerPhe......AsnAsnIleTyrLeuLysLysAsnIl 438

    387 TGGGTTTGAATCAATCATATGTTTTGCTCTATCACCTTGAC......... 347
        |:::||| |||::::::   :::|||      :::   :::
    438 eAsnPhe.AsnLysThrThrIleLeuThrIleAsnLysAsnThrPheLys 454

    346 ...TGTTCCATGAACCATAGTATCCTGCTATAATACCTTGTTTCCTTTTT 300
           |||:::::::::::::::|||:::
    455 IleCysAsnIleThrGlnAsnIleIle...................... 463

    299 TTATACTCCTCTAAAATTTGTCTTGGTGATTTTCTATCATTTTTATTTAT 250
         |||   |||:::::::::::::|||   ::::::   ::::::      :
    464 .TyrIleProPheAsnTyrLeuLeuSerPheIleGluAsnLeuIleProP 480

    249 ATTTTTATTATTTTTTGTAATTCTAGAGTTACACCCATCTCCGCAATAAG 200
        ::::::::|||                          :::|||||||
    480 heIleHisTyr.................AsnAspSerIleArgAsnLeu 490

    199 ATCCATATTCTACAAAATTACTATTATTCGATTTAAGATGTGAAAAAGAC 150
        :::   |||   ::::::                              ||
    491 MetSerIleLysMetHis...........................Th 497

    149 TCATGTATTTCCGTTTGATGAGTTTTATTTTTATTTTCCCGTATTATTCC 100
        |::::::   |||              :::   :::|||   |||
    497 rGlnIleValPro..............IleIleTyrProAsnLeuSerA 509

     99 CAATGAATTATTTATATTATTTTTTCCTTTCAAGGTTCTGGAAT 56
          :::        |||   |||:::|||:::::::::   |||
    509 snIleIleThrAsnTyrAsnPheIleLeuAsnLysTyrLeuAsn 523
```

```
seq_name: SwissProt_39:YXDM_BACSU

seq_documentation_block:
ID   YXDM_BACSU     STANDARD;      PRT;   622 AA.
AC   P42424;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 70.5 KDA PROTEIN IN IDH 3'REGION.
GN   YXDM OR B65G.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=97021444; PubMed=8867804;
RA   Yoshida K.-I., Fujimyra M., Yanai N., Fujita Y.;
RT   "Cloning and sequencing of a 23-kb region of the Bacillus subtilis
RT   genome between the iol and hut operons.";
RL   DNA Res. 2:295-301(1995).
RN   [2]
RP   SEQUENCE OF 1-404 FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=95039891; PubMed=7952181;
RA   Yoshida K.-I., Sano H., Miwa Y., Ogasawara N., Fujita Y.;
RT   "Cloning and nucleotide sequencing of a 15 kb region of the Bacillus
RT   subtilis genome containing the iol operon.";
RL   Microbiology 140:2289-2298(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D45912; BAA08316.1; -.
DR   EMBL; D14399; BAA03303.1; -.
DR   EMBL; Z99124; CAB15999.1; -.
DR   SubtiList; BG11129; yxdM.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     20     40       POTENTIAL.
FT   TRANSMEM     56     76       POTENTIAL.
FT   TRANSMEM    118    138       POTENTIAL.
FT   TRANSMEM    154    174       POTENTIAL.
FT   TRANSMEM    195    215       POTENTIAL.
FT   TRANSMEM    219    239       POTENTIAL.
FT   TRANSMEM    279    299       POTENTIAL.
FT   TRANSMEM    498    518       POTENTIAL.
FT   TRANSMEM    558    578       POTENTIAL.
FT   TRANSMEM    590    610       POTENTIAL.
SQ   SEQUENCE   622 AA;  70528 MW;  B2FAFAD30D27370E CRC64;


alignment_scores:
            Quality:  121.00              Length:    397
              Ratio:  0.602                Gaps:     15
 Percent Similarity:  50.630   Percent Identity: 19.899

alignment_block:
 US-09-579-383-2 x YXDM_BACSU   ..

 Align seg 1/1  to: YXDM_BACSU  from: 1  to: 622

    505 TATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGA 554
        ||||||||||:::|||:::          :::::::::   ::::::|||
    211 TyrGlyMetValLeuLys......GlyAsnValHisGlyAlaGluProPh 225

    555 TGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAG 604
         :::|||:::||||||:::::   |||   ||||||:::::      :
    225 eIleIleLeuLeuLeuThrValIleGly...ThrTyrPhePhePheSerG 241

    605 AAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTT 654
        :::::::::|||   ::::::   :::   ||||||      :::   |||
    241 lnSerSerIle...TrpIleLeuArgAlaLeuLysLysTrpLysThrPhe 256

    655 GATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTTACAA 704
           |||   |||   :::|||   |||
    257 TyrLeuArgGlyLysAsnIleIleTrp..................... 265

    705 CTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGT 754
                                                :::   :::|||:
    266 ................................ValSerAspLeuV 270

    755 TAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCA 804
        ::   :::          :::   :::|||:::   |||   :::
    270 alTyrArgLeuLysAspAsnAlaArgLeuPhePheIleValSerIleIle 286

    805 AATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGAT.. 852
        :::|||:::          ::::::||||||   ::::::   |||:::
    287 SerAlaValAlaPheThrAlaThrGlyValLeuAlaMetTyrLysSerTh 303

    853 ..........GAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAA 892
                   |||||||||   |||:::   ::::::||||||   :::
    303 rValGlyAlaGluGluSerAlaTyrGluMetGluTyrLeuSerTyrSerA 320
```

```
 893 TAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGA 942
       :::   :::::::::   |||                   :::
 320 snAsnProLysGluGlnThrHisLeuLysAspIleAspHisGluLeuLys 336

 943 ACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATT 992
     |||       :::   |||:::   :::|||:::|||||||||:::   ::
 337 Thr......HisGlyPheThrTyrThrLysAspLysIleAspValSerTy 351

 993 TATTCAAACATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATA 1042
     :::::::       :::   ||||||:::   |||      |||::::::
 351 rValArgTyrGlnGluGlyGluThrValProProValTyrMetIleSerG 368

1043 TGTACTTATCCCATTTATATTTT.............GGTTTAAAATAT 1077
           :::      ||||||              |||||||||
 368 luSerAspAlaAlaLysTyrPheHisValLysValAsnGlyLeuLysGlu 384

1078 AACATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGATT 1127
     :::   ::::::   :::       :::   ::::::|||      :::
 385 AspGluAlaValTyrPheProGlyThrTyrAspArgAsnPheLysAsnGl 401

1128 TAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATG 1177
      :::|||:::::::   :::|||||:::       |||:::        :
 401 uAlaProAspGlnLeuLysLeuLeuAsnGlnLysGlyGluLeuSerAspG 418

1178 ATAAAAATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTA 1227
     ::|||                                          |||
 418 lnLys.....................................Leu 420

1228 TTTATGAAAGAA...CAATTACCAACTGGATCATTCGATGTAGATATTTT 1274
        :::|||||||   |||   |||       |||:::::::::   |||
 421 SerValLysGluValGlnLysProLeuIleSerLeuAsnAlaIleIleAl 437

1275 TCTTACAAAT..................................... 1284
      :::::::::
 437 aValAsnAspGlnThrPheAspGlnLeuLysSerLeuGlyAspLysAlaS 454

1285 ..............ATTTGGAAACATTTAAATCCTGAAGTACAAACTCCA 1320
                     :::   |||   :::   :::::::::
 454 erLeuTyrGlyTyrSerTyrAspHisTrpLysAspSerLeuGluIleSer 470

1321 AAAGACCTTACT...................ATAACTGAAAACCCTGA 1349
     :::::::|||                     |||      :::   ::
 471 GlnSerLeuGlnAsnGluIleTyrGlyAsnTyrIleAspValHisSerAs 487

1350 AGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCA 1399
     :   :::|||           |||         :::   :::|||::::
 487 pPheAlaSerLysAlaGlyThrTyrTyrAspThrValGlnLeuProSerL 504

1400 TAGGGATATAT.........TACAAACACAATGATGCTATATGGAAAACT 1440
     ::::::::::         :::      |||      :::   :::
 504 euSerLeuPheIleGlyLeuPheIleGlyAsnCysIleLeuCysArgGly 520

1441 AGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTT 1490
        |||:::||| ::::::::::|||      ||||||::::::
 521 ThrSerPheSer.TyrPheArgLeuPhe...ThrAspLeuAspGluAspA 536

1491 GGTCAAAGTAT...GCTATGAAAAAATATGCGATGGGAAAGCAGCCCATT 1537
       ::::::|||   ::::::   |||   :::::::::::::::   :::
 536 rgGluArgTyrArgSerLeuAlaLysIleGlyLeuSerGluArgGluMet 552

1538 ATTATAACACTGACTATAAAGAAAGCTCTATTATTATAT 1576
          ::::::||||||:::   |||:::|||::::::
 553 AlaGlnSerValThrIleGlnLeuAlaIleLeuPhePhe 565

seq_name: SwissProt_39:CHIT_NPVOP

seq_documentation_block:
ID   CHIT_NPVOP     STANDARD;      PRT;   550 AA.
AC   O10363;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=164623;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97271300; PubMed=9126251;
RA   Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA   Rohrmann G.F.;
RT   "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT   polyhedrosis virus genome.";
RL   Virology 229:381-399(1997).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U75930; AAC59123.1; -.
DR   HSSP; P07254; 1CTN.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR000886; ER_target.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS00014; ER_TARGET; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;
KW   Endoplasmic reticulum.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    550       PROBABLE ENDOCHITINASE.
FT   ACT_SITE    304    304       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    146    146       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    172    172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    344    344       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        547    550       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   550 AA;  60733 MW;  77947F5CF00E07BD CRC64;

alignment_scores:
            Quality:  120.00          Length:    292
              Ratio:  0.882             Gaps:     13
 Percent Similarity:  46.575   Percent Identity: 21.233

alignment_block:
 US-09-579-383-2 x CHIT_NPVOP   ..

 Align seg 1/1  to: CHIT_NPVOP  from: 1  to: 550

     319 ATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACA 368
         :::|||:::|||:::   :::|||:::      |||
     149 ValAlaAlaTyrPheValGluTrpGlyValTyrGly......ArgGlyPh 163

     369 TATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTC 418
           :::|||:::   |||:::   ::::::   ::::::   ::::::
     163 eProValAspLysValProLeuProAsnLeuSerHisLeuLeuTyrGlyP 180

     419 GCATTAATATG.............TTATATGAT.............. 438
           |||   :::               |||   |||
     180 heIleProIleCysGlyGlyAspGlyLeuAsnAspAlaLeuLysThrIle 196

     439 ................GTATCTAGACCATTTAATGGAAGACAAAGATT 470
                         :::::::|||      :::||||||      ||
```

```
     197 ProGlySerPheGluAlaLeuGlnArgSerCysLysGlyArgAlaAspPh 213

     471 CCTATTAAGAAAACAC.............................G 487
         |   :::      |||                              |
     213 eLysValAlaIleHisAspProTrpAlaAlaIleGlnLysProGlnLysG 230

     488 GCTTAGAA......TATGAAACCTATGGTATGATGCTTAATGAAATTAGA 531
         ||:::::::         |||   |||          ::::::::::::
     230 lyValSerAlaTrpAsnGluProTyrLysGlyAsnPheGlyGlnLeuMet 246

     532 CGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGG 581
               :::   :::::::|||   :::   ||||||   |||:::|||||
     247 AlaAlaLysLeuAlaAsnProHisLeuLysIleLeuProSerIleGlyGl 263

     582 AGAAACC..................TATATGATAGATATAGAAAAAGAAA 613
         |   |||                    :::|||   |||   :::|||:::
     263 yTrpThrLeuSerAspProPheTyrPheMetHisAspAlaAspLysArgA 280

     614 TTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGAT...TTA 660
               :::|||::::::::::   ::::::::::::   :::      :::
     280 rgValPheValGluSerValLysGluPheLeuGlnValTrpLysPhePhe 296

     661 GATGGTGTAGATATTGACTGGGAA...CCACATGGGAAGTTTTACAACTT 707
         ||||||||||||||||||||||||   |||   ||||||      |||
     297 AspGlyValAspIleAspTrpGluPheProGlyGlyLysGlyAlaAsnPr 313

     708 A.........AATGAATTAAATTTTTCAAATTATTATATTAAATTAATTA 748
                 |||      :::   ::::::   |||:::   |||::::
     313 oAlaLeuGlyAsnGlyGluArgAspAlaAspThrTyrLeuValLeuLeuL 330

     749 ACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGT 798
         ::   ||||||      :::   |||
     330 ysGluLeuArgAlaMetLeuAspGlu..................... 338

     799 TCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAA 848

     338 ..................................................... 338

     849 AGATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAA 898
                                                 |||:::::::|
     339 .................................LeuGlnLeuGlnT 343

     899 CAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTT 948
         ||||||||      :::   ::::::|||:::   |||:::
     343 hrAsnLysThrTyrGluLeuThrSerAlaIleSerSerGlyTyrAspLys 359

     949 ATTAATATT......TTTAATACAGCAAAGGAGAAAATAGATCTTGTATT 992
         |||   :::      :::::::::|||::::::   :::|||   :::||
     360 IleAlaValValLysTyrAspAlaAlaGlnArgPheLeuAspLysIlePh 376

     993 TATTCAAACATACAATTTAGAA......ACTACAAATCCAGATATAATGG 1036
         |:::::::::|||::::::::::         :::|||   |||:::
     376 eLeuMetSerTyrAspPheLysGlyAlaTrpSerAsnThrAspLeu.... 391

    1037 TAGATATGTACTTATCCCATTTATAT 1062
                |||   :::   ||||||
     392 .....GlyTyrGlnThrThrLeuTyr 398

seq_name: SwissProt_39:HEMK_RICPR

seq_documentation_block:
ID   HEMK_RICPR     STANDARD;      PRT;   518 AA.
AC   Q9ZCB3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HEMK PROTEIN HOMOLOG.
GN   HEMK OR RP847.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- FUNCTION: PROBABLE METHYLTRANSFERASE.
CC   -!- SIMILARITY: IN THE N-TERMINAL REGION; BELONGS TO THE HEMK FAMILY
CC       OF MODIFICATION METHYLASES.
CC   -!- SIMILARITY: IN THE C-TERMINAL REGION; BELONGS TO THE UPF0155
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235273; CAA15271.1; -.
DR   InterPro; IPR003358; Methyltransf_4.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR000818; TEA.
DR   Pfam; PF02390; Methyltransf_4; 1.
DR   SMART; SM00426; TEA; 1.
DR   PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW   Transferase; Methyltransferase; Complete proteome.
FT   DOMAIN        1    300       HEMK.
FT   DOMAIN      301    518       UPF0155.
SQ   SEQUENCE   518 AA;  59564 MW;  C66BB8DB273704A1 CRC64;


alignment_scores:
            Quality:  119.00              Length:    422
              Ratio:  0.583                 Gaps:     25
 Percent Similarity:  48.341   Percent Identity: 21.564

alignment_block:
 US-09-579-383-2 x HEMK_RICPR   ..

 Align seg 1/1  to: HEMK_RICPR  from: 1  to: 518

      97 TTGGGAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATACA 146
         :::|||::::::      :::::::|||:::   |||:::
     107 IleGlyLeuVal...ValSerArgAsnAsnLeuHisMet........... 118

     147 TGAGTCTTTTTCACATCTTAAATCGAATAATAGT................ 180
               ||||||   |||||||||   :::|||
     119 .......PheSerLysLeuLysSerLeuAspSerValLeuThrThrGlnS 133

     181 .....AATTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGT...... 219
             |||::::::::|||   |||:::      |||:::||||||
     133 erTyrAsnIleLeuGluLeuGlyThr......GlySerGlyCysIleAla 147

     220 .......................AACTCTAGAATTACAAAAAATAATAA 245
                                 |||:::::::|||      ::::::
     148 IleSerLeuLeuCysGluLeuProAsnThrAsnIleIleAlaThrAspIl 164

     246 AAATATAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGT 295
          :::::::::                 :::   ::::::|||      :::
     164 eSerValAspAlaIleLysValAlaLysSerAsnSerIleLysTyrAsnV 181

     296 ATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAAC 345
               |||   |||   |||            :::::::::|||
     181 alThrAspArgIleGlnIleIle............HisSerAsnTrpPhe 193

     346 AGTCAAGGTGATAGAGCAAAA...CATATGATTGATTCAAACCCA..... 387
         ::::::   |||:::   |||      :::|||   |||||||||
```

```
 194 GluLysLeuAspLysGlnLysPheAspPheIleValSerAsnProProTy 210

 388 .......ATGGTGTCAATTTTATATATTGCATTTGCTCGCATTAATATGT 430
             ::::::   |||   :::|||:::      ||||||
 210 rIleSerHisThrGluLysLeuLysMetAlaIleGluThrIleAsn.... 225

 431 TATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGA 480
       |||:::   |||                          ::::::   :::
 226 ..TyrGluProSer.................IleAlaLeuPheAlaGlu 235

 481 AAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATTAG 530
     :::   ||||||      |||:::|||:::::::::   ::::::   ::
 236 GluAspGlyLeu......GluAlaTyrSerIleIleAlaLysAsnAlaLy 250

 531 ACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTG 580
     :::::::      :::::::|||:::      |||:::|||::::::|||
 250 sGlnPhe......LeuLysProAsnGlyLysIleIleLeuGluIleGly. 264

 581 GAGAAACCTATATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAA 630
                           ::::::       |||:::|||
 265 ......................PheSerGlnAlaAlaLysValSerLys 273

 631 ATATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTG 680
     |||:::   |||         :::::::::|||   :::      |||
 274 IlePhe...LeuAsnTyrGlyTyrAsnIleAspTyrIleTyrArgAspLe 289

 681 GGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATT 730
      :::   |||::::::      :::::::::   :::|||:::::::::
 289 uGlnSerHisAsnArgValIleGluIleSerProIleAsnLeuAsnArgS 306

 731 ATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCG......GAA 774
       |||   :::   |||:::          |||::::::          :::
 306 erTyrAlaArgArgIleGly.........LysSerLeuSerLysMetGln 319

 775 GAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGT 824
     :::||||||:::
 320 GlnLysLeuLeu..................................... 323

 825 TTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTA 874
                                |||:::|||   |||         |
 324 .........................AspAsnGluLeuPro.........L 329

 875 AATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCA 924
     ||:::|||   ::::::   :::   :::::::|||   :::::::::
 329 ysTyrLeuPheSerLysGluLysPheLysSerGluLysArgLysValPhe 345

 925 GCGATGTTATCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGA 974
           ::::::   |||         :::::::::|||   ||||||
 346 LeuGluIleGlyPheGlyMetGlyGluHisLeuIleAsnGlnAlaLysIl 362

 975 GAAAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAATC 1024
      :::   |||   :::||||||
 362 eAsnProAspThrLeuPheIleGly..................... 370

1025 CAGATATAATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAA 1074
               |||::::::||||||:::
 371 ...........ValGluValTyrLeuAsn.................... 376

1075 TATAACATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGG 1124
        :::::::::   :::|||      |||   :::||||||      ::
 377 ...GlyValAlaAsnValLeuLysHisSerAlaGlnHisAsnIleThrAs 392

1125 ATTT......AGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAA 1168
     :|||         |||:::|||   :::|||:::
 392 nPheLeuLeuPheProAsnAsnLeuAspLeuIle............... 403

1169 CAATACATGATAAAAATCAAAATAATAATAGGGCAGATGGTATAGGGATA 1218
       ::::::|||         ||||||      |||||||||   |||
 404 ..LeuAsnAsp......LeuProAsnAsnSerLeuAspGlyIleTyrIle 417

1219 ..............TGG.......................CATTTATT 1229
                   |||                       ::::::||
 418 LeuPheProAspProTrpIleLysAsnLysLysLysLysLysArgIlePh 434

    1230 TATGAAAGAACAATTA 1245
         |   ||||||:::|||
     434 eAsnLysGluArgLeu 439

seq_name: SwissProt_39:Y875_METJA

seq_documentation_block:
ID   Y875_METJA     STANDARD;      PRT;   748 AA.
AC   Q58285;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0875.
GN   MJ0875.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67531; AAB98882.1; -.
DR   HSSP; P08622; 1XBL.
DR   TIGR; MJ0875; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   748 AA;  90731 MW;  5515ED3B541562EA CRC64;


alignment_scores:
            Quality:  116.50              Length:    642
              Ratio:  0.445                Gaps:     34
 Percent Similarity:  40.810   Percent Identity: 20.249

alignment_block:
 US-09-579-383-2 x Y875_METJA   ..

 Align seg 1/1  to: Y875_METJA  from: 1  to: 748

     241 AATAAAAATATAAATAAAAATGATAGAAAATCACCA.............. 276
         ||||||      :::|||||||   :::|||:::
     193 AsnLysPheAsnAspLysAsnTyrLysLysAlaGluIlePheAsnLysHi 209

     277 ...................AGACAAATTTTAGAGGAGTATAAAAAAAGGA 307
                               |||   ::::|||:::   :::|||   |
     209 sPheTyrTyrAsnLeuMetAlaGlnLysPheGluSerGluArgLysPheL 226

     308 AACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGAT 357
         ||::::::      |||   ||||||         :::::::::||||||
     226 ysGluAla......AlaGluTyrTyr.........LysLysSerGlyAsp 237

     358 AGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATAT 407
```

```
          |||:::   ||||||:::            :::|||   |||
  238 ThrIleLysGlu...IleAspGlu............LysIleAlaTyrAs 249

  408 TGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATG 457
        :::|||:::                  :::::::::   :::|||
  249 pGluTyrAlaAsnSerTyrLysTrpLeuAlaIleGluAsnLysTyrAsn. 265

  458 GAAGACAAAGATTC.........................CTATTAAGA 480
        ::::::::::|||                          :::   :::
  266 ..LysGluLysPheGluGluTyrIleAsnLysAlaIleGluPheSerGlu 281

  481 AAACACGGC.........TTAGAATATGAAACCTATGGTATGATGCTTAA 521
      |||:::|||          ||||||        |||:::      ::
  282 LysArgGlyAspLysLeuGlnGluTyrTyrTyrLeuGlyLeuLysTyrAs 298

  522 TGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTAT 571
      :::::::   |||
  298 pHisLeuValArg..................................... 302

  572 CCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAAGAAATTGATTAT 621
                    :::      |||:::|||::::::|||||||||
  303 .................PheAlaAsnAspLeuGluGluLysIleAspTyr 313

  622 GTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGA 671
      :::   |||      :::
  314 IleLysLysSerLysGluTyr............................. 320

  672 TATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAATT 721
                        :::|||:::   ::::::::::::|
  321 .......................TyrTyrArgSerLysSerPheGluP 329

  722 TTTCAAATTAT..................................... 732
      ||:::::::|||
  329 heAlaLysTyrMetGluTyrLeuGluTyrTyrTyrGlnPheLysTyrGlu 345

  733 ..............TATATTAAATTAATTAACTTGTTAAGAAAAACTAT 767
                  |||   |||   :::|||:::|||:::|||:::
  346 LeuLeuAsnGlyAsnTyrGluLysAlaLeuAsnPheLeuGlnLysAlaLy 362

  768 TCCGGAAGAAAAGTTAATTTCAATTTCTGGT......TCATCAAATGCTG 811
        :::   |||   ::::::||||||:::        |||:::   :
  362 sLysSerLeuLysAsnValLysIleSerAsnIleIlePheSerLysTyrT 379

  812 CATTATCATGCGTTTCAGGAGTTGCATCTTTC................... 843
      ::|||:::|||   :::   ::::::   |||
  379 hrLeuGluCysAspGluLeuIleCysArgPheTyrLeuSerIleSerGln 395

  844 .......................TGTAAAGATGAAGAATCTCCATATAA 869
                                    ||||||          ::
  396 GlyGluPheLysLysSerValGluLeuLeuAspGluTyrLeuGluIleSe 412

  870 CACTAAATTTTTGTCTGAACAAATAGAAACAAATAAA.............. 906
      :   |||:::||||||:::       :::|||:::|||
  412 rLeuLysIleLeuSerAspTrpLysAsnThrArgLysTyrLysPheTyrG 429

  907 ..GAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAAT 954
          |||      ::::::   :::||||||      :::|||
  429 luTyrLeuLysProCysValGluIleLeuSerLysGluSerPhe...... 443

  955 ATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATA 1004
                          |||   ||||||   :::::::::
  444 ..................ThrLysAspAspLeuPheLeuLeuGluAspVa 454

 1005 CAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTAC....... 1047
          ||||||:::   |||      |||   :::::::::|||
  454 lIleLeuGluValArgAsnGluLysIleSerLeuAsnLeuTyrLysIleC 471

 1048 .....TTATCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATA 1092
          |||::::::   |||   :::|||      |||
  471 ysSerLeuThrTyrAlaTyrValSerLeuTrpLysAsn............ 483

 1093 TTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAA 1142
                                          :::   ::::::||
```

```
  484 .................................LysIleArgAspLy 488

 1143 AGAATTATTAGAATTGGTAGGA............AAAACAATACATGATA 1180
      ||||:::||||||   :::              |||   |||      :
  488 sGluIleLeuGluLysIleLysLeuLysIleIleLysGlnIleThrThrG 505

 1181 AAAATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTT 1230
      :::::   :::::::::      :::   |||   |||     :::
  505 luAspValSerLysAspLeuGluAsnArgIleLysIle..GlnMetAlaI 521

 1231 ATGAAAGAACAATTACCAACTGGATCATTCGA.................. 1262
        ||||||      ::::::|||:::::::|||
  521 leGluArg......AsnAspTrpLeuLeuArgLeuProProAlaLeuVal 535

 1263 ..............TGTAGATATTTTTCTTACAAATATTTGGAAACATT 1297
                     |||   ||||||      :::   |||:::   ||
  536 GluAsnPheAspAsnCysIleTyrPheLeuGluGluValLeuAspGluPh 552

 1298 TAAATCCTGAAGT..................................... 1310
      ||||   :::
  552 eLysHisAlaAlaTyrArgGluPheTyrArgLeuLeuGluAsnTyrLeuL 569

 1310 ..................................................... 1310

  569 ysIleIleValGluPheAsnAlaMetAlaLeuTrpGlyAspThrTrpLys 585

 1311 ....................ACAAACTCCAAAAGACCTTAC........ 1331
                          :::|||   |||:::|||:::
  586 GlnIleLeuGluGluLysIleSerAsnArgLysLysProPheGluIlePh 602

 1332 .....................................TATAACTGAAAAC 1344
                                          ||| |||:::
  602 eThrPheGlyAspPheAlaGlnSerLeuArgLeuLeuLysAsn.GluGly 618

 1345 CCTGAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCC 1394
        |||   |||:::::::||||||:::   :::      |||:::
  619 CysGluTyrCysLysAsnIleAspAspGluIlePheAspLeuLeuAspLy 635

 1395 AACCATAGGGATATATTACAAACACAATGATGCTATATGGAAAACTAGAT 1444
          :::   |||      ::::::|||               :::::::
  635 sHisValLysIle......ArgAsnAsn............LeuSerHisA 646

 1445 CTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTC 1494
      ::::::::      ||||||                ::::   |||   :::
  646 spPheAsnAlaLysAlaProGluAspIleAspIleLysGlyAspThrLeu 662

 1495 AAAGTATGCTATGAA..............AAAATATGCGATGGGAAAGC 1529
      |||:::   |||:::                    ||||||      |||::
  663 LysIleIleTyrSerThrLeuAsnAlaPheProIleCys...ValLysVa 678

 1530 AGCCCATTATTATAACACTGACTATAAAGAAAGCTCTATTATTATATGGA 1579
      ::::|||             |||   |||:::                |||
  678 lThrHis............AspLysLysLysPro............TrpT 687

 1580 AAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCTCCGGAAGGT 1629
        |||         :::         |||:::|||      |||:::
  687 yrGlyValGluValIle.........TrpAsnGln...LeuProLysLys 699

 1630 CAGGCACTAGAGTCATACACAAAACTAGATGCATCCAAATGTCCAGGGAT 1679
      ::::::|||   |||   |||:::|||      :::           ||
  700 MetSerLeuTyrSerAsnThrGluLeuLysLysGlyCysLeuTyrTyrIl 716

 1680 AGAAGAATGG.................AATAAAAAATATCCACAT.... 1707
      ||||   :::                  ||||||   :::|||
  716 eGluProTyrSerAsnIleLeuAspLysAsnLysMetHisProLysIleI 733

 1708 ..AAACCACTAGAAGTAGAGGAA 1728
          |||:::||||||:::::
  733 leIleProIleGluValSerLys 740
```

seq_name: SwissProt_39:RPOP_AGABT

seq_documentation_block:

```
ID   RPOP_AGABT     STANDARD;      PRT;  1102 AA.
AC   P33539;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6).
OS   Agaricus bitorquis.
OG   Mitochondrion.
OG   Plasmid pEM.
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Agaricales; Agaricaceae; Agaricus.
OX   NCBI_TaxID=5343;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 24666 / AG4;
RX   MEDLINE=91347410; PubMed=1879001;
RA   Robison M.M., Royer J.C., Horgen P.A.;
RT   "Homology between mitochondrial DNA of Agaricus bisporus and an
RT   internal portion of a linear mitochondrial plasmid of Agaricus
RT   bitorquis.";
RL   Curr. Genet. 19:495-502(1991).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63075; CAA44799.1; -.
DR   PIR; S28104; S28104.
DR   InterPro; IPR002092; RNA_pol_phage.
DR   Pfam; PF00940; RNA_pol; 1.
DR   PROSITE; PS00489; RNA_POL_PHAGE_2; 1.
DR   PROSITE; PS00900; RNA_POL_PHAGE_1; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase;
KW   Mitochondrion; Plasmid.
FT   ACT_SITE    734    734       BY SIMILARITY.
FT   ACT_SITE    804    804       BY SIMILARITY.
FT   ACT_SITE    980    980       BY SIMILARITY.
SQ   SEQUENCE   1102 AA;  126664 MW;  F72FDD308D1AB5B4 CRC64;


alignment_scores:
            Quality:  116.50              Length:    463
              Ratio:  0.537                Gaps:     21
 Percent Similarity:  46.868   Percent Identity: 21.166

alignment_block:
 US-09-579-383-2 x RPOP_AGABT    ..

 Align seg 1/1  to: RPOP_AGABT  from: 1  to: 1102

      10 AAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTCTGCAAATTC 59
         |||   |||:::         :::   :::|||   |||:::::::::::
     184 LysGluSerValValAlaLysValGluGluIleAsnTyrGluValHisAs 200

      60 CAGAACCTTGAAAGGAAAAAATAATATAAAT..................... 90
         ::::   :::||||||::::::|||   |||
     200 nLysPheIleLysGlyAsnGlyAsnThrAsnPheSerAsnArgAsnLeuT 217

      91 ............................AATTCATTGGGAATAATA... 108
                                     |||::::::   |||:::
     217 hrGluIleLysSerIleLeuLysGluLeuAsnLysMetGluIleLeuAsp 233

     109 ...CGGGAAAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTT 155
            |||   ||||||   :::|||   ::::::::::::   :::   ::
     234 AsnArgIleAsnLysLeuSerThrLysGluSerAspLeuLeuLysValIl 250

     156 TTCACATCTTAAATCGAATAAT........................... 177
         ::::::::::   ::::::|||
     250 eLysGluIleLeuAspSerAsnLeuIleIleGluAspLysGlnLeuAlaI 267

     178 .....AGTAATTTTGTAGAATATGGA...TCTTATTGCGGAGATGGGTGT 219
               ::::::   |||||||||      ::::::          :::
     267 leGluLysThrValValGluTyrGluLeuThrPhePheArgHisAsnMet 283

     220 AACTCTAGAATTACAAAAAATAAT......AAAAATATAAATAAAAATGA 263
         :::::::::   |||:::|||:::         ||||||      :::
     284 AspThrHisGluThrArgAsnLysIleIleHisAsnIleTyrProLysLe 300

     264 TAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAA............. 300
          :::|||:::      ::::::|||   :::||||||
     300 uAsnLysAlaTyrThrGluLeuLeuAlaAsnTyrLysLeuAsnArgTyrS 317

     301 ..AAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGT 348
           |||   |||:::::|||         :::      |||   :::|||
     317 erLysIleLysLysSerIle.........HisLeuIleSerAsnLysSer 330

     349 CAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAAT 398
         :::|||   ::::::|||:::||||||          ::::::   ||
     331 GluGlyThrLysSerLysGluMetIleLysLeuIleValValLeuValIl 347

     399 TTTATATATT.......................GCATTTGCTCGCATTA 424
         ||||||||||                      :::|||   :::|||
     347 eLeuTyrIleGlyIleAspLysCysIleSerTyrSerPheTyrGlnIleI 364

     425 ATATGTTATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTA 474
               |||         :::|||      ::::::|||   ::::::
     364 leAsnLeuLeuThrAsnAlaArgAspGlyThrSerArgThrAsnIleAla 380

     475 TTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGATG........ 516
         :::         :::|||   ::::::        ::::::
     381 Ile.........AsnLeuGlyPheArgIleIleLysValLeuLysTyrIl 394

     517 ......................CTTAATGAAATTAGACGTATCAGAAAAG 544
                               ||||||   :::      |||:::|||:
     394 eLysLeuAspGluAsnProSerLeuAsnAlaLeuTyrProIleAsnLysL 411

     545 TACGTCCAGATGTAATTATTCTTTTATCCTTAGGT............GGA 582
         :::::   ::::::      |||   :::   |||            |||
     411 euLysAspGluIleSerLysLeuAspAsnGluGlyIleTyrTrpIleGly 427

     583 GAAACCTATATGATAGATATAGAAAAAGAAATTGATTATGTGGATAAAAT 632
         :::|||   :::      |||      :::   |||   |||   :::
     428 AspThrLeuLeuGlyLeuIleThrAlaAsnCysAspIleValValGluGl 444

     633 ATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGG 682
          ||||||      :::               |||   |||   :::   |
     444 uLeuLysTrpAsnSer..............GlyLysAspSerGlnLeuG 456

     683 AACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTAT 732
         ||                  ::::::|||:::      :::||||||
     456 lu..............ValArgIleAsnAspLysPheIleSerAsnLeu 467

     733 TATATTAAATTAATTAACTTGTTAAGA...AAAACTATTCCGGAAGAAAA 779
            ::::::   ||||||:::::::::        :::   |||   ::
     468 ThrValSerGlyIleAsnIleValGlnLeuProMetLeuThrGluProAr 484

     780 GTTAATTTCAATTTCTGGT.....................TCATCAAATG 808
         :   ||||||   :::|||                      |||:::::::
     484 gLysIleSerSerAspGlyLeuTyrPheProTyrIleAsnSerAspThrT 501

     809 CTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAA 858
         ::   |||         :::|||                      :::
     501 hrAsnLeuHisLeuPheGluGlyGluLeuIle..............Lys 512

     859 TCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGA 908
```

```
     :::   ||||||         |||   ::::::   |||:::::::::
 513 GlyLysTyrAsn.........LeuArgAspHisThrGluAlaSerGluMe 526

 909 ATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATATTT 958
      |||:::   :::                              ||||||   :
 526 tLeuTyrSerSer.........................IleAsnTyrL 534

 959 TTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATACAAT 1008
     ::|||:::   |||   ||||||:::   ::::::::::::
 534 euAsnSerIleLysPheLysIleAsnLysAlaMetLeuAsnPheIleLeu 550

1009 TTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTATCCCATTT 1058
        |||      |||   |||                       |||   ||
 551 AlaGluTrpAspAsnLysAsp.....................SerLysLe 560

1059 ATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTCATTAG 1108
     |:::   |||      ||||||:::
 560 uPheLysGly......TyrAsnMetLeu..................... 567

1109 AACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTAGAATTG 1158
                                   |||   :::|||||||
 568 ...............................LysProIleLeuGluThr 573

1159 GTAGGAAAAACAATACATGATAAAAATCAAAATAATAAT 1197
        :::|||   |||   :::::::::   ::::::|||
 574 AspSerLysGluIleLysGluGluLysValSerSerAsn 586
```

```
OM of: US-09-579-383-2 to: SPTREMBL_17:*   out_format : pfs

Date: Mar 21, 2002  4:52 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09579383/runat_20032002_151240_29414/app_query.fasta_1.1849
 -DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09579383_@CGN1_1_172 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-579-383-2
Query length: 1764
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 143.370000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
sp_invertebrate:Q9NJX9 + 3130.00 3643.02  2.1e-195  587  ! Q9njx9 plasmodium gallinaceum
sp_invertebrate:Q9U4I9 +  635.00  740.42  1.5e-33   378  ! Q9u4i9 plasmodium falciparum
sp_invertebrate:Q9NJX8 +  597.00  697.47  4.4e-31   323  ! Q9njx8 plasmodium falciparum.
sp_bacteria:P96168     +  163.50  185.70  0.0062    729  ! P96168 vibrio harveyi. chitinas
sp_organelle:Q34192    -  158.50  181.65  0.0129    590  ! Q34192 crithidia oncopelti. nad
sp_invertebrate:O97230 +  154.00  169.11  0.0270   1410  ! O97230 plasmodium falciparum
sp_invertebrate:O97291 +  150.50  162.34  0.0466   1946  ! O97291 plasmodium falciparum
sp_invertebrate:O77393 +  150.50  160.51  0.0474   2423  ! O77393 plasmodium falciparum
sp_fungi:Q06350        +  149.00  171.78  0.0529    511  ! Q06350 saccharomyces cerevisiae
sp_bacteria:O54328     +  148.50  170.39  0.0574    563  ! O54328 enterobacter sp. chitina
sp_invertebrate:Q25802 -  148.50  165.92  0.0597    960  ! Q25802 plasmodium falciparum.
sp_archea:Q9UWR7       +  148.00  163.37  0.0654   1215  ! Q9uwr7 pyrococcus kodakaraensi
sp_bacteria:Q9ALZ0     +  147.50  169.22  0.0666    563  ! Q9alz0 serratia liquefaciens. e
sp_bacteria:O83008     +  144.50  165.73  0.1043    563  ! O83008 serratia marcescens. chi
sp_invertebrate:O96206 +  144.00  155.34  0.1224   1817  ! O96206 plasmodium falciparum
sp_invertebrate:Q21861 +  143.00  165.97  0.1282    444  ! Q21861 caenorhabditis elegans
sp_invertebrate:Q9BJY0 +  142.50  150.12  0.1579   2752  ! Q9bjy0 plasmodium yoelii yoe
sp_bacteria:P97034     +  142.00  162.83  0.1515    562  ! P97034 enterobacter agglomerans
sp_invertebrate:O77380 +  142.00  149.19  0.1707   2870  ! O77380 plasmodium falciparum
sp_bacteria:Q59326     +  141.00  162.95  0.1739    482  ! Q59326 clostridium thermocellum
sp_invertebrate:O97275 +  141.00  151.87  0.1916   1812  ! O97275 plasmodium falciparum
sp_bacteria:O07088     +  140.50  160.06  0.1912    635  ! O07088 bacillus thuringiensis.
sp_virus:Q9EMP3        +  140.50  154.48  0.2008   1238  ! Q9emp3 amsacta moorei entomopo
sp_organelle:Q9G9H3    +  140.00  152.55  0.2189   1453  ! Q9g9h3 schizophyllum commune (
sp_virus:Q9WFU7        +  139.50  159.80  0.2203    570  ! Q9wfu7 helicoverpa armigera nuc
sp_invertebrate:Q9TY98 +  139.50  148.15  0.2440   2295  ! Q9ty98 plasmodium falciparum
sp_virus:Q9PYU0        +  139.00  158.88  0.2381    594  ! Q9pyu0 xestia c-nigrum granulos
sp_invertebrate:Q9BK46 +  139.00  144.97  0.2689   3130  ! Q9bk46 plasmodium falciparum
sp_invertebrate:Q9BK45 +  139.00  144.64  0.2697   3254  ! Q9bk45 plasmodium falciparum
sp_bacteria:Q9RH34     +  138.50  158.77  0.2555    561  ! Q9rh34 serratia marcescens. chi
sp_virus:Q99H06        +  138.50  158.64  0.2558    570  ! Q99h06 heliocoverpa armigera nu
sp_bacteria:Q9EY54     +  138.50  156.58  0.2604    729  ! Q9ey54 vibrio sp. fi:7. chitina
sp_bacteria:Q9ZIX4     +  138.50  153.49  0.2676   1054  ! Q9zix4 pseudoalteromonas sp. s
sp_invertebrate:O96204 +  138.00  141.81  0.3178   3973  ! O96204 plasmodium falciparum
sp_virus:Q995W8        +  137.50  157.47  0.2970    570  ! Q995w8 helicoverpa armigera nuc
sp_invertebrate:Q9U5A3 +  137.00  152.20  0.3335    999  ! Q9u5a3 plasmodium falciparum
sp_invertebrate:Q9U0K8 +  136.50  145.97  0.3775   1960  ! Q9u0k8 plasmodium falciparum
sp_invertebrate:O16237 +  136.50  145.70  0.3784   2025  ! O16237 caenorhabditis elegan
sp_invertebrate:Q9NGK8 +  136.00  156.76  0.3683    504  ! Q9ngk8 wuchereria bancrofti.
sp_virus:O92482        +  135.50  155.41  0.3995    552  ! O92482 bombyx mori nuclear poly
sp_invertebrate:O96223 +  135.50  142.77  0.4462   2500  ! O96223 plasmodium falcipa
sp_organelle:Q33559    -  135.00  156.67  0.4236    443  ! Q33559 leishmania tarentolae
sp_organelle:Q9XMS2    -  134.50  146.54  0.4962   1386  ! Q9xms2 tetrahymena pyriform
sp_invertebrate:O77347 +  133.50  149.61  0.5552    836  ! O77347 plasmodium falcipar
sp_bacteria:Q9RMB8     +  133.50  149.18  0.5573    880  ! Q9rmb8 arthrobacter sp. chit

seq_name: sp_invertebrate:Q9NJX9

seq_documentation_block:
ID   Q9NJX9      PRELIMINARY;      PRT;   587 AA.
AC   Q9NJX9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   Plasmodium gallinaceum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5849;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20209408; PubMed=10744721;
RA   Vinetz J.M., Valenzuela J.G., Specht C.A., Aravind L., Langer R.C.,
RA   Ribeiro J.M.C., Kaslow D.C.;
RT   "Chitinases of the Avian Malaria Parasite Plasmodium gallinaceum, a
RT   Class of Enzymes Necessary for Parasite Invasion of the Mosquito
RT   Midgut.";
RL   J. Biol. Chem. 275:10331-10341(2000).
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; AF064079; AAF63208.1; -.
DR   InterPro; IPR001579; Chitinase_2.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Glycosidase; Hydrolase.
SQ   SEQUENCE   587 AA;  67918 MW;  23D1ABA87DEBEE37 CRC64;


alignment_scores:
           Quality: 3130.00              Length:    587
             Ratio:  5.332                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-579-383-2 x Q9NJX9   ..

 Align seg 1/1  to: Q9NJX9  from: 1  to: 587

       1 ATGAATTTTAAAATATCAATATTTTTAATTATAGTATCCATCTTGTATTC 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetAsnPheLysIleSerIlePheLeuIleIleValSerIleLeuTyrSe 17

      51 TGCAAATTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTCATTGG 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 rAlaAsnSerArgThrLeuLysGlyLysAsnAsnIleAsnAsnSerLeuG 34

     101 GAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATACATGAG 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lyIleIleArgGluAsnLysAsnLysThrHisGlnThrGluIleHisGlu 50

     151 TCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATC 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 SerPheSerHisLeuLysSerAsnAsnSerAsnPheValGluTyrGlySe 67

     201 TTATTGCGGAGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rTyrCysGlyAspGlyCysAsnSerArgIleThrLysAsnAsnLysAsnI 84

     251 TAAATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 leAsnLysAsnAspArgLysSerProArgGlnIleLeuGluGluTyrLys 100

     301 AAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 LysArgLysGlnGlyIleIleAlaGlyTyrTyrGlySerTrpAsnSerGl 117
```

```
 351 AGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTT 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 nGlyAspArgAlaLysHisMetIleAspSerAsnProMetValSerIleL 134

 401 TATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCA 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 euTyrIleAlaPheAlaArgIleAsnMetLeuTyrAspValSerArgPro 150

 451 TTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGA 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 PheAsnGlyArgGlnArgPheLeuLeuArgLysHisGlyLeuGluTyrGl 167

 501 AACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTC 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 uThrTyrGlyMetMetLeuAsnGluIleArgArgIleArgLysValArgP 184

 551 CAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGAT 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 roAspValIleIleLeuLeuSerLeuGlyGlyGluThrTyrMetIleAsp 200

 601 ATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGA 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 IleGluLysGluIleAspTyrValAspLysIleLeuLysLeuValAsnAs 217

 651 TTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTT 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 pPheAspLeuAspGlyValAspIleAspTrpGluProHisGlyLysPheT 234

 701 ACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAAC 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 yrAsnLeuAsnGluLeuAsnPheSerAsnTyrTyrIleLysLeuIleAsn 250

 751 TTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTC 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuLeuArgLysThrIleProGluGluLysLeuIleSerIleSerGlySe 267

 801 ATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAG 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 rSerAsnAlaAlaLeuSerCysValSerGlyValAlaSerPheCysLysA 284

 851 ATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 spGluGluSerProTyrAsnThrLysPheLeuSerGluGlnIleGluThr 300

 901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTAT 950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 AsnLysGluLeuHisArgAlaAlaAlaMetLeuSerAlaGlyThrPheIl 317

 951 TAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 eAsnIlePheAsnThrAlaLysGluLysIleAspLeuValPheIleGlnT 334

1001 CATACAATTTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTA 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 hrTyrAsnLeuGluThrThrAsnProAspIleMetValAspMetTyrLeu 350

1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 SerHisLeuTyrPheGlyLeuLysTyrAsnIleThrIleIleLeuGlyPh 367

1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTAT 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 eSerLeuGluHisAsnArgGlyGlyPheSerProGluAsnLysGluLeuL 384

1151 TAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAATAGG 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 euGluLeuValGlyLysThrIleHisAspLysAsnGlnAsnAsnAsnArg 400

1201 GCAGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAAC 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 AlaAspGlyIleGlyIleTrpHisLeuPheMetLysGluGlnLeuProTh 417

1251 TGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGGAAACATTTAA 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 417 rGlySerPheAspValAspIlePheLeuThrAsnIleTrpLysHisLeuA 434

1301 ATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAA 1350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 434 snProGluValGlnThrProLysAspLeuThrIleThrGluAsnProGlu 450

1351 GACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCAT 1400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 451 AspCysSerThrIleAspGluTyrValProGlyLeuValIleProThrIl 467

1401 AGGGATATATTACAAACACAATGATGCTATATGGAAAACTAGATCTTATT 1450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 467 eGlyIleTyrTyrLysHisAsnAspAlaIleTrpLysThrArgSerTyrS 484

1451 CAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAAAGTA 1500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 484 erIleHisAlaProGlyValAspArgTyrGluTrpAspLeuValLysVal 500

1501 TGCTATGAAAAAATATGCGATGGGAAAGCAGCCCATTATTATAACACTGA 1550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 501 CysTyrGluLysIleCysAspGlyLysAlaAlaHisTyrTyrAsnThrAs 517

1551 CTATAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTAATTA 1600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 517 pTyrLysGluSerSerIleIleIleTrpLysGlyGluProTyrLeuIleL 534

1601 AATGGTGGCAACAAGGACCTCCGGAAGGTCAGGCACTAGAGTCATACACA 1650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 534 ysTrpTrpGlnGlnGlyProProGluGlyGlnAlaLeuGluSerTyrThr 550

1651 AAACTAGATGCATCCAAATGTCCAGGGATAGAAGAATGGAATAAAAAATA 1700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 551 LysLeuAspAlaSerLysCysProGlyIleGluGluTrpAsnLysLysTy 567

1701 TCCACATAAACCACTAGAAGTAGAGGAACAATATGAACAAGAAGTGGATT 1750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 567 rProHisLysProLeuGluValGluGluGlnTyrGluGlnGluValAspL 584

1751 TACCATTACAA 1761
     ||||||||||||
 584 euProLeuGln 587
```

seq_name: sp_invertebrate:Q9U4I9

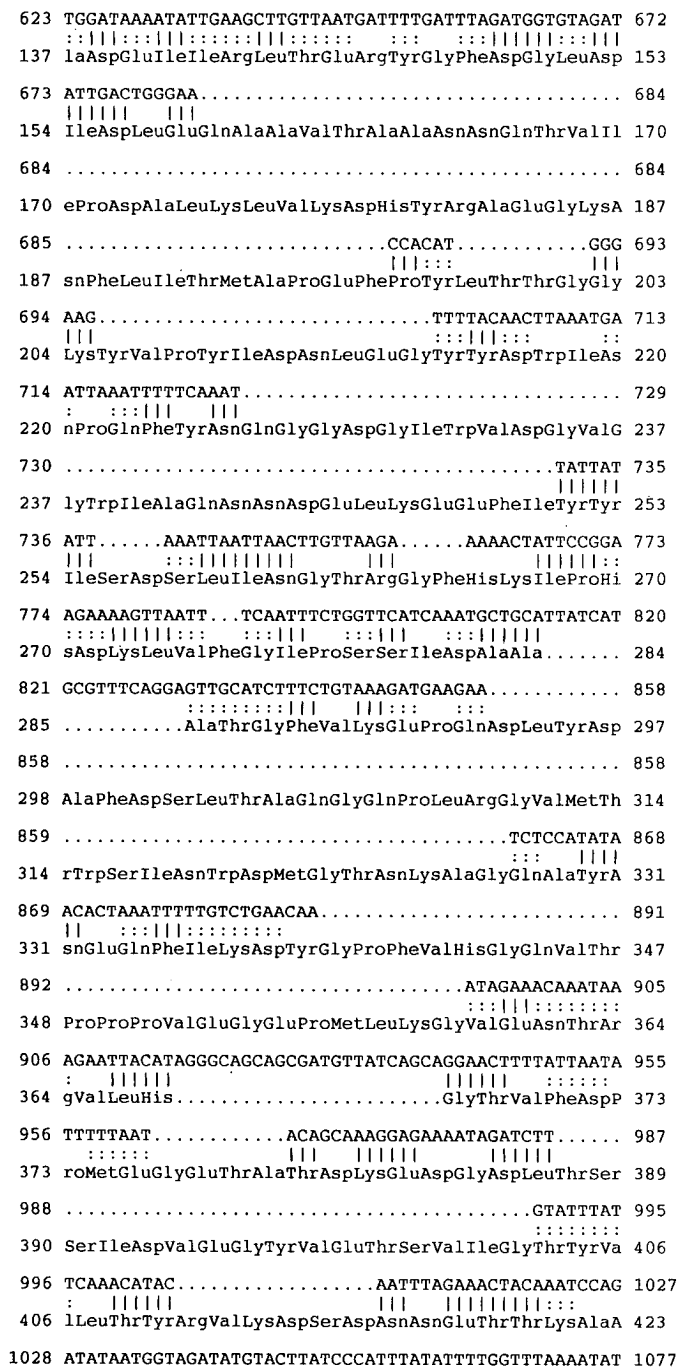
seq_documentation_block:

ID Q9U4I9 PRELIMINARY; PRT; 378 AA.
AC Q9U4I9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHITINASE.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=20040676; PubMed=10570198;
RA Vinetz J.M., Dave S.K., Specht C.A., Brameld K.A., Kuntz I.D., Xu B.,
RA Hayward R., Fidock D.A.;
RT "The chitinase PfCHT1 from the human malaria parasite Plasmodium
RT falciparum lacks proenzyme and chitin-binding domains and displays
RT unique substrate preferences.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14061-14066(1999).
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF172445; AAF16902.1; -.
DR InterPro; IPR001579; Chitinase_2.
DR InterPro; IPR002114; PTS_HPr_ser.
DR PROSITE; PS01095; CHITINASE_18; 1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.

```
KW   Glycosidase; Hydrolase.
SQ   SEQUENCE   378 AA;  42786 MW;  5DF86F2AC93C81E3 CRC64;


alignment_scores:
            Quality:  635.00              Length:    345
              Ratio:  2.560                Gaps:      5
 Percent Similarity:  71.884   Percent Identity: 37.391

alignment_block:
 US-09-579-383-2 x Q9U4I9   ..

 Align seg 1/1  to: Q9U4I9  from: 1  to: 378

     265 AGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAAGG 314
          |||||||:::|||||||:::|||:::::   :::|||:::      :::||
      37 ArgLysAsnProArgGluIleIleLysThrPheLysGluSerGlyLysGl 53

     315 TATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAA 364
          |||||||   |||||||||   ||||||   |||   :::   :::   |
      53 yIleIleGlnGlyTyrTyrProSerTrpValSerTyrAsnHisAsnLeuL 70

     365 AACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTT 414
          ||         |||   ||||||   :::::::::::::::::::::|||
      70 ys.........AspLeuAsnProAsnLeuAsnValValHisMetSerPhe 83

     415 GCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGAAGACA 464
          |||::::::::::::   ||||||      :::   :::   |||
      84 AlaLysMetAspLeuSerTyrAspSerIleGluSerIleValGlySerPr 100

     465 AAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGA 514
              ::::::            ||||||||||||          :::
     100 oLeuLeuPheLysSerLeuIleGlyLeuGluTyrIleGlyLeuAsnGluT 117

     515 TGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATT 564
            :::|||:::      ::::::||||||:::||||||||||:::|||:::
     117 yrPheAsnAspAlaMetAsnLeuArgLysAlaArgProAspIleIleMet 133

     565 CTTTTATCCTTAGGTGGAGAAACCTAT...ATGATAGATATAGAAAAAGA 611
          ||||||||||||||||||||||||||         ::::::::::::
     134 LeuLeuSerLeuGlyGlyGluThrTyrHisProSerSerPheAspSerAl 150

     612 AATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAG 661
          ::::::   |||:::||||||   :::||||||:::::::::   :::|
     150 aLeuAsnAlaValGluLysIleAlaAsnLeuValAspGluLeuGlyPheA 167

     662 ATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTTACAACTTAAAT 711
          |||||:::|||:::|||:::||||||:::|||:::|||   :::||||||
     167 spGlyIleAspValAspTyrGluProAsnGlySerPheAspGlyLeuAsn 183

     712 GAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAA 761
          :::   :::   :::::::::::::::::   ::::::   ||||||::
     184 AspLysGluLysAlaAspPhePheValGlnTyrValThrLysLeuArgGl 200

     762 AACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTG 811
          :   :::   ::::::||||||||||||||||||   |||||||||:::|
     200 uTyrMetCysAspAspLysLeuIleSerIleSerGlnSerSerAsnGlyA 217

     812 CATTATCATGCGTT......TCAGGAGTTGCATCTTTCTGTAAAGATGAA 855
          ||||||||||:::       :::         ::::::|||   |||:::
     217 laLeuSerCysIleGlyPheAsnAspProLysLysIleCysMetAspAsp 233

     856 GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAA 905
          |||:::||||||||||:::|||:::::::::::::      :::::::::||
     234 GluAlaProTyrAsnSerLysTyrPheAsnLysPro...AspValLysLy 249

     906 AGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATA 955
          |||||||   |||||||||   |||   |||||||||      |||   :
     249 sGluLeuLeuArgAlaAlaGlnMetAlaSerAlaGlyGlyAlaIleTyrL 266

     956 TTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATAC 1005
          :::::|||:::   |||:::   |||||||:::||||||:::|||||||:::
     266 euMetAsnAsnLeuLysAspMetIleAspMetValPheValGlnThrPhe 282

    1006 AAT...TTAGAAACTACAAATCCAGATATAATGGTAGATATGTACTTATC 1052
          |||      ::::::|||:::      :::|||   :::::::|||   ||
     283 AsnTyrThrAsnSerThrAspSerThrValMetLysGluLeuTyrAspSe 299

    1053 CCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTT 1102
          |:::   |||:::|||   ||||||:::   :::|||||||:::|||||||:
     299 rTyrAlaTyrTyrGlyLysLysTyrAspTyrValIleIleMetGlyPheT 316

    1103 CATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTATTA 1152
          ::|||                  |||:::|||::::::|||   |||:::
     316 hrLeuMetPheProSerThrProPheAsnProAsnAspLysMetLeuVal 332

    1153 GAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAATAGGGC 1202
          :::   :::|||      :::      :::|||:::   |||:::|||||
     333 LysSerIleGlyAspPheValLysThrGluAsnLysLeuAsnLysArgAl 349

    1203 AGATGGTATAGGGATATGGCATTTATTTATGAAAGAACAATTACCAACTG 1252
          |||||||:::|||:::|||   |||         :::         :::
     349 aAspGlyPheGlyLeuTrpSerLeuSerSerAspAsnAlaAlaHisAsnG 366

    1253 GATCATTCGATGTAGATATTTTTCTTACAAATATT 1287
            ::::::   ::::::   |||:::   ::::::
     366 luGlnLeuAlaIleGluTyrPheValGluSerLeu 377

seq_name: sp_invertebrate:Q9NJX8

seq_documentation_block:
ID   Q9NJX8      PRELIMINARY;      PRT;   323 AA.
AC   Q9NJX8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CHITINASE (FRAGMENT).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20209408; PubMed=10744721;
RA   Vinetz J.M., Valenzuela J.G., Specht C.A., Aravind L., Langer R.C.,
RA   Ribeiro J.M.C., Kaslow D.C.;
RT   "Chitinases of the Avian Malaria Parasite Plasmodium gallinaceum, a
RT   Class of Enzymes Necessary for Parasite Invasion of the Mosquito
RT   Midgut.";
RL   J. Biol. Chem. 275:10331-10341(2000).
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; AF072442; AAF63209.1; -.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR002114; PTS_HPr_ser.
DR   PROSITE; PS01095; CHITINASE_18; 1.
DR   PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
KW   Glycosidase; Hydrolase.
FT   NON_TER       1      1
FT   NON_TER     323    323
SQ   SEQUENCE   323 AA;  36502 MW;  9079ADEFE0F754D2 CRC64;


alignment_scores:
            Quality:  597.00              Length:    310
              Ratio:  2.642                Gaps:      5
 Percent Similarity:  72.903   Percent Identity: 39.355

alignment_block:
 US-09-579-383-2 x Q9NJX8   ..

 Align seg 1/1  to: Q9NJX8  from: 1  to: 323

     310 CAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGGTGATAG 359
          :::||||||||||   |||||||||   ||||||   |||   :::   ::
       1 LysGlyIleIleGlnGlyTyrTyrProSerTrpValSerTyrAsnHisAs 17
```

```
 360 AGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTATATATTG 409
     :   |||         |||   ||||||   ::::::::::::::::::
  17 nLeuLys.........AspLeuAsnProAsnLeuAsnValValHisMetS 31

 410 CATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGA 459
     ::||||||:::::::::::::   ||||||      :::   :::   |||
  31 erPheAlaLysMetAspLeuSerTyrAspSerIleGluSerIleValGly 47

 460 AGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGG 509
             ::::::           ||||||||||||          ::
  48 SerProLeuLeuPheLysSerLeuIleGlyLeuGluTyrIleGlyLeuAs 64

 510 TATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAA 559
     :      :::|||:::      ::::::||||||:::|||||||||:::|
  64 nGluTyrPheAsnAspAlaMetAsnLeuArgLysAlaArgProAspIleI 81

 560 TTATTCTTTTATCCTTAGGTGGAGAAACCTAT...ATGATAGATATAGAA 606
     ||:::|||||||||||||||||||||||||||        :::::::::
  81 leMetLeuLeuSerLeuGlyGlyGluThrTyrHisProSerSerPheAsp 97

 607 AAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGA 656
     :::   ::::::   |||:::||||||   :::|||||||:::::::::
  98 SerAlaLeuAsnAlaValGluLysIleAlaAsnLeuValAspGluLeuGl 114

 657 TTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTTACAACT 706
      :::||||||:::|||:::|||:::||||||:::|||:::|||   :::|
 114 yPheAspGlyIleAspValAspTyrGluProAsnGlySerPheAspGlyL 131

 707 TAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTTA 756
     ||||||:::   :::   ::::::::::::::::::   ::::::   |||
 131 euAsnAspLysGluLysAlaAspPhePheValGlnTyrValThrLysLeu 147

 757 AGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAA 806
     |||:::   :::   ::::::|||||||||||||||||   ||||||||
 148 ArgGluTyrMetCysAspAspLysLeuIleSerIleSerGlnSerSerAs 164

 807 TGCTGCATTATCATGCGTT......TCAGGAGTTGCATCTTTCTGTAAAG 850
     |:::|||||||||||:::      :::         :::::::|||   |
 164 nGlyAlaLeuSerCysIleGlyPheAsnAspProLysLysIleCysMetA 181

 851 ATGAAGAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACA 900
     ||:::|||:::|||||||||:::|||:::::::::::::      ::::::
 181 spAspGluAlaProTyrAsnSerLysTyrPheAsnLysPro...AspVal 196

 901 AATAAAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTAT 950
     :::||||||||   |||||||||   |||   |||||||||      ||
 197 LysLysGluLeuLeuArgAlaAlaGlnMetAlaSerAlaGlyGlyAlaIl 213

 951 TAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAA 1000
     |   :::::::|||:::   |||:::   ||||||:::|||||||:::||||
 213 eTyrLeuMetAsnAsnLeuLysAspMetIleAspMetValPheValGlnT 230

1001 CATACAAT...TTAGAAACTACAAATCCAGATATAATGGTAGATATGTAC 1047
     ||:::|||      ::::::|||:::      :::|||   :::::::|||
 230 hrPheAsnTyrThrAsnSerThrAspSerThrValMetLysGluLeuTyr 246

1048 TTATCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGG 1097
        |||:::   |||:::|||   ||||||:::   :::|||||:::||
 247 AspSerTyrAlaTyrTyrGlyLysLysTyrAspTyrValIleIleMetGl 263

1098 TTTTTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAAT 1147
     ||||:::|||                  |||:::|||:::::::|||   |
 263 yPheThrLeuMetPheProSerThrProPheAsnProAsnAspLysMetL 280

1148 TATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAATAAT 1197
     ||:::::::   :::|||      :::      :::|||:::   |||:::
 280 euValLysSerIleGlyAspPheValLysThrGluAsnLysLeuAsnLys 296

1198 AGGGCAGATGGTATAGGGATATGGCATTTA 1227
     ||||||||||||:::|||:::|||   |||
 297 ArgAlaAspGlyPheGlyLeuTrpSerLeu 306

seq_name: sp_bacteria:P96168
```

```
seq_documentation_block:
ID   P96168      PRELIMINARY;      PRT;   729 AA.
AC   P96168;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CHITINASE A.
GN   CHIA.
OS   Vibrio harveyi.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=669;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BB7;
RX   MEDLINE=98274739; PubMed=9611805;
RA   Svitil A.L., Kirchman D.L.;
RT   "A chitin-binding domain in a marine bacterial chitinase and other
RT   microbial chitinases: implications for the ecology and evolution of
RT   1,4-beta-glycanases.";
RL   Microbiology 144:1299-1308(1998).
DR   EMBL; U81496; AAC46383.1; -.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001638; SBP_bac_3.
DR   InterPro; IPR003610; Chitin_bind3.
DR   PROSITE; PS01095; CHITINASE_18; UNKNOWN_1.
DR   PROSITE; PS01039; SBP_BACTERIAL_3; 1.
DR   SMART; SM00495; ChtBD3; 2.
SQ   SEQUENCE   729 AA;  79016 MW;  77213CE43DB95A07 CRC64;


alignment_scores:
            Quality:  163.50            Length:    610
              Ratio:  0.678               Gaps:     32
 Percent Similarity:  39.508   Percent Identity: 19.836

alignment_block:
 US-09-579-383-2 x P96168   ..

 Align seg 1/1  to: P96168  from: 1  to: 729

     253 AATAAAAATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAAA 302
         |||   :::|||      :::|||
      25 AsnGlySerAspMetThrAsnPro..................... 32

     303 AAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGG..... 342
              :::|||:::::::::||||||:::   :::|||
      33 ....AspSerGlyValValValGlyTyrTrpGlnAsnTrpCysAspGlyG 48

     343 ..AACAGTCAAGGTGATAGAGCAAAA......CATATGATTGATTCAAAC 384
           :::   ||||||   :::|||           :::   :::   |||
      48 lyGlyTyrGlnGlyGlyAsnAlaProCysValThrLeuAspGluValAsn 64

     385 CCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATA 434
         ||||||   :::|||:::   :::::::|||   ::::::         ||
      65 ProMetTyrAsnIleValAsnValSerPheMetLysVal.........Ty 78

     435 TGATGTATCTAGACCATTTAATGGAAGA...CAAAGATTCCTATTAAGAA 481
         |||||||:::        :::||||||         |||   |||
      78 rAspValAla.........AspGlyArgIleProThrPheLysLeuAspP 92

     482 AACAC...GGCTTAGAATATGAAACCTATGGTATGATGCTTAATGAAATT 528
               ||||||:::   |||   :::      :::::::::|||:::
      92 roThrValGlyLeuSerGluGluGlnPheIleAspGlnValSerGluLeu 108

     529 AGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGG 578
         :::::      :::                  :::||||||:::|||||
     109 AsnLysGlnGlyArg..............SerValLeuLeuAlaLeuGl 120

     579 TGGAGAAACCTATATGATAGATATAGAA......AAAGAAATTGATTATG 622
         ||||          :::::::::|||         |||      ::::
     120 yGlyAlaAspAlaHisValGluLeuGluThrGlyAspGluArgAlaPheA 137
```



seq_name: sp_organelle:Q34192

seq_documentation_block:
ID Q34192 PRELIMINARY; PRT; 590 AA.
AC Q34192;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 5.
GN ND5.
OS Crithidia oncopelti.
OG Mitochondrion.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
OX NCBI_TaxID=5657;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-068;
RA Dmitriy Maslov A.;
RL Submitted (OCT-1990) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S-068;
RA Maslov D.A., Horvath A., Gwang II K., Kolesnikov A.A.;
RL Submitted (AUG-1991) to the EMBL/GenBank/DDBJ databases.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; X56015; CAA39492.1; -.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.

```
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE   590 AA;  71650 MW;  1A2FE619E5B0C02E CRC64;


alignment_scores:
            Quality:  158.50              Length:    681
              Ratio:  0.576                 Gaps:     41
 Percent Similarity:  40.382   Percent Identity: 21.145

alignment_block:
 US-09-579-383-2/rev x Q34192   ..

 Align seg 1/1  to: Q34192  from: 1  to: 590

    1743 TTCTTGTTCATATTGTTCCTCTACTTCTAGTGGTTTATGTGGATATTTTT 1694
         |||:::|||:::|||||||:::   |||   :::|||:::      |||||
       4 PhePhePheMetLeuPhePheLeuPheGlyPhePheLeuGlyThrPhePh 20

    1693 T..............ATTCCATTCTTCTATCCC..............  1675
         |               :::   |||:::
      20 eLeuGlyArgHisLeuLeuSerPheTrpLeuSerIleLeuMetThrValP 37

    1674 ..................TGGACATTTGGATGCATCTAGTTTTGTGTA 1645
                           ::::::|||   ||||||      |||:::
      37 heLeuValMetValThrMetPheSerPhePheCysIleSerValCysLeu 53

    1644 TGACTCTAGTGC...CTGACCTTCCGGAGGTCCTTGTTGCCACCATTTAA 1598
                 |||         |||         |||:::      :::
      54 TyrGlyTyrCysTyrTyrAspPheCysLeuIleLeuMetLeuAspLeuCy 70

    1597 TTAAATATGGTTCCCCTTTCCATATAATAATAGAGCTTTCTTTATAGTCA 1548
             |||::::::   |||      |||:::      :::||||||:::
      70 sPheIlePheMetSerPheTyrCysAsnGlyPheTyrLeuPheIleLeuP 87

    1547 GTGTTATA...ATAATGGGCTGCTTTCCCATCGCATAT...TTTTTCATA 1504
           :::|||   ::::::   ||||||   :::   |||   |||:::
      87 heLeuIleAspLeuValPheCysPheIleLeuPheTyrAlaPheTyrTyr 103

    1503 GCATACTTTGACCAAGTC......CCATTCATATCTGTC.........TA 1469
           |||:::|||   :::         |||:::::::::         ::
     104 MetTyrTyrAspLeuMetLeuLysArgPhePheAsnIlePheTrpTrpPh 120

    1468 CACCAGGTGCATGAATTGAATAAGATCTAGTTTTCCATATAGCATCATTG 1419
         ::::   |||||||||   :::      |||:::   |||
     120 eValLeuCysMetAsnPhePheIleLeuSerTyrAspTyr.......... 133

    1418 TGTTTGTAATATATCCCTAT...GGTTGGAATAACGAGTCCTGGAACATA 1372
                 :::      |||   ||||||      |||      ::::::
     134 ........LeuThrAlaTyrCysGlyTrp......GluLeuLeuGlyLeu 145

    1371 TTCATCTATTGTGCTACAGTCTTCAGGGTTTTCAGTTATAGTAAGGTCTT 1322
         |||   :::            |||   :::      :::          ||
     146 PheSerPhe............PheLeuIleSerTyrPheTrpTyrArgPh 158

    1321 TTGGAGTTTGTACTTCAGGATTTAAATGTTTCCAAATATTTGTAAGAAAA 1272
         |::::::|||   ||| |||||||||   |||      |||:::   |||
     158 ePheAlaLeuLysPhe.GlyPheLysSerPhe......PheIleSerLys 172

    1271 ATATCTACATCGAATGATCCAGTTGGTAATTGTTCTTTCATAAATAAATG 1222
         |||:::       ::::::   |||   :::|||::::::   |||::::
     173 IleGlyAspVal.LeuLeuLeuLeuSerPheValMetThrPheIleSerT 189

    1221 CC............ATATCCCTATACCATCTGCCCTATTATTATTTTGAT 1184
         ::            |||:::::::|||   :::   :::         |||
     189 hrGlyTyrGlyMetIleAsnPheTyrPheValAsnPheLeuCysValAsp 205

    1183 TTTTATCATGTATTGTTTTTCCTACCAATTCTAATAATTCTTTATTTTCG 1135
         ||||||   :::   |||:::|||   :::||||||:::      |||
     206 PheTyrPheThrAlaPheMetLeuPheLeuLeuIleMetCysAlaPheTh 222

    1134 ...............GGACTAAATCCACCTCTGTTATGTTCTAATGAAA 1100
                         ||||||:::      |||      :::   |||
     222 rLysSerThrGlnPheGlyLeuHisIleTrpLeuProAspAlaMetGluG 239

    1099 AACCT............................AATATGATTGTG... 1083
           |||                           :::::::::|||
     239 lyProIleProValSerAlaLeuIleHisAlaAlaThrLeuValValCys 255

    1082 .....................................ATGTT 1078
                                             :::
     256 GlyIleLeuLeuValSerPhePhePheTrpCysPheAspPheTrpLeuAl 272

    1077 ATATTTTAAACCAAAATATAAATGGGATAAGTACATATCTACCATTATAT 1028
           ||||||   |||         |||::::::   |||   :::::::::|
     272 aTyrPheTyrProLeuIleGlyTrpSerSerLeuIleLeuValMetMetS 289

    1027 CTGGATTTGT..............................AGTTTCT 1011
         ||  :::|||                             :::|||
     289 er..LeuCysValPheTyrAsnPheAspAlaLysArgPheValAlaPheS 305

    1010 AAATTGTATGTTTGAATAAATACAAGATCTATTTTCTCCTTTGC...... 967
               :::|||      :::::::::::::::      ||||||
     305 erThrIleCysGlnIleSerPheSerMetPheCysCysLeuCysLeuAsp 321

     966 ............TGTATTAAAAATATTAATAAAAGTTCCTGCTGATAACA 929
                       |||:::
     322 LeuTyrValGlyCysLeu............................... 327

     928 TCGCTGCTGCCCTATGTAATTCTTTATTTGTTTCTATTTGTTCAGACAAA 879
                               |||:::|||::::::::::|||   :::
     328 .....................PhePheCysTyrHisMetPheTyrLysA 337

     878 AA......TTTAGTGTT......ATATGG..................... 862
                 |||   |||      :::|||
     337 laThrLeuPheIleValLeuGlyValTrpIleHisLeuPhePheGlyLeu 353

     861 ..............AGATTCTTCATCTTTACAGAAAGATGC........ 835
                       |||||||||:::       |||
     354 GlnAspValArgCysTyrPhePheIleTyrPheCysGlyCysValLeuAl 370

     834 ............................AACTCCTGAAACGCATGATAAT 813
                                     ||||||               |
     370 aArgMetLeuLeuValPheAlaLeuLeuAsnSer..............C 382

     812 GCAGCATTTGATGAACCAGAAATTGAAATTAACTTTTCTTCCGGAATAGT 763
         |||||
     382 ysSer.......................................... 383

     762 TTTTCTTAACAAGTTAATTAATTTAATATAATAATTTGAAAAATTTAATT 713

     383 ..................................................  383

     712 CATTTAAGTTGTAAAACTTCCCATGTGGTTCCCAGTCAATATCTACACCA 663
                               :::|||||||      :::::::|||
     384 .....................LeuTrpPheLeuCysGlyPheTyrCysL 393

     662 TCTAAATCAAAATCATTAACAAGCTTCAATATTTTATCCAC...ATAATC 616
               :::   :::   :::   :::          :::|||   |||:::
     393 ysAspLeuLeuLeuCysThrLeuMetLeuValSerPheHisPheIleLeu 409

     615 AATTTCTTTTTCTAT...ATCTATCATATAGGTTTCTCCACCTAAGGATA 569
         :::|||:::|||      |||:::   |||   |||::::::
     410 GluPheLeuPheValCysIlePhePheIlePhePheThrVal........ 423

     568 AAAGAATAATTACATCTGGACGTACTTTTCTGATACGTCTAATTTCATTA 519
                   |||:::::::||| ||||||:::   :::::
     424 .............IleTyrAsnTyr.PheLeuLeuPhePheLeu...... 433

     518 AGCATCATACCATAGGTTTCATATTCTAAGCCGTGTTTTCTTAATAGGAA 469
                                         ||||||:::   :::
     434 ................................CysPheValPheLysCy 439

     468 TCTTTGTCTTCCATTAAATGGTCTAGATACATCATATAACATATTAATGC 419
          :::||||||
     439 sPheCysLeu..................................... 442
```

```
     418 GAGCAAATGCAATATATAAAATTGACACCATTGGGTTTGAATCA..... 375
     443 ..AlaAspCysLeuPheLeuLeu.....PheAspPheGluCysCysLeu 456

     374 ATCATATGTTTTGCTCTATCACCTTGACTGTTCCATGAACCATAGTATCC 325
     457 IleTyrCysPhe.LeuGlyLeuTyrMetCys.....PheIleL 469

     324 TGCTATAATACCTTGTTTCCTTTTTTTATACTCCTCTAAAATTTGTCTTG 275
     469 euIlePhePheValIleAspPheLeuTyrIlePheValPheSerSerTyr 485

     274 GTGATTTTCTATCATTTTTATTTATATTTTTATTATTTTTTGTAATTCTA 225
     486 CysMetPheTrpSerPheTyrLeuTyrTyrPheAsnPhePheAspIleAl 502

     224 GAGTTACACCCATCTCCGCAATAAGATCCATATTCTACAAAATTACTATT 175
     502 aSerPheThr..................................... 505

     174 ATTCGATTTAAGATGTGAAAAAGACTCATGTATTTCCGTTTGATGAGTTT 125
     506 .....ValPheValMetIleSerLeuAlaPhe 514

     124 TATTTTTATTTTCCCGTATTATTCCCAATGAATTATTTATATTATTTTTT 75
     515 IleTyrTyrGlyCysValLeuPhe.....TyrPhePh 525

      74 C.....CTTTCAAGGTTCTGGAATTTG 53
     525 eAsnValAspCysIleMetLeuPheTrpArgIle 536
```

```
seq_name: sp_invertebrate:O97230

seq_documentation_block:
ID   O97230      PRELIMINARY;      PRT;  1410 AA.
AC   O97230;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL PROTEIN, PFC0195W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=99376085; PubMed=10448855;
RA   Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA   Churcher C.M., Craig A., Davies R.M, Devlin K., Feltwell T.,
RA   Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA   Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA   Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA   Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA   Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT   "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT   falciparum .";
RL   Nature 400:532-538(1999).
DR   EMBL; AL034558; CAB38995.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1410 AA;  170544 MW;  A1590A8CB8011EB3 CRC64;


alignment_scores:
           Quality:  154.00            Length:    651
             Ratio:  0.526               Gaps:     30
 Percent Similarity:  45.008   Percent Identity: 20.123

alignment_block:
 US-09-579-383-2 x O97230   ..

 Align seg 1/1  to: O97230  from: 1  to: 1410
```

```
      31 ATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGGAAAAAA 80
      86 IleLysGlnIleArgHisSerAsnAsnLysAspAsnSerAsnAsnAsnAs 102

      81 TAATATAAATAATTCATTGGGAATAATACGGGAAAATAAAAATAAAACTC 130
     102 nAsnAsnAsnAsn.....AsnAsnAsnSerAsnA 112

     131 ATCAAACGGAAATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGT 180
     112 snAsnAsnAsnAsnAsnAsnSerAsnAsnAsnAsnAsnSerAsnAsnAsn 128

     181 AATTTTGTAGAATATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAAT 230
     129 Asn.....AsnSerAsnAs 133

     231 TACAAAAAATAATAAAAATATAAATAAAAATGATAGAAAATCACCAAGAC 280
     133 nAsnAsnAsnAsnSerAsnLysAsnSerAsnAsnAsnAsnAsn..... 147

     281 AAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATTATA......GCA 324
     148 .....TyrAsnAspTyrTyrLysAspGlyGlnAsnIleValIleAsnThr 162

     325 GGATACTATGGTTCATGGAACAGTCAAGGTGATAGAGCAAAACATATGAT 374
     163 AsnTyrThrSerSerTyrTyrLysHisLysLysSerValAsnAsnIlePh 179

     375 TGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTA 424
     179 eIleAsnAsnAspValPheSerSerLeuGlnAsnIlePheLeuLeuIleA 196

     425 ATATGTTA.....TATGATGTATCTAGACCATTTAATGGA 459
     196 snLysIleAsnLysAsnPhePheGlnArgValGlnProLysPheAspGly 212

     460 AGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAAACCTATGG 509
     213 AsnGlnLysGlnIleIle..... 218

     510 TATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAA 559
     219 .....IleAspAsnValValTyrIleAsnLys...HisProTyrValL 232

     560 TTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATATAGAAAAA 609
     232 euHisIleLeu...LeuLysAsnAsnAsnValAsnIleAsnIleTyrLys 247

     610 GAAATTGATTATGTG.....GATAAAATATTGAAGCTTGTTAATGA 650
     248 GluIleAspIleLeuLeuAlaSerGluTyrPheSerHisLeuCysAsn.. 263

     651 TTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTTTT 700

     263 ..... 263

     701 ACAACTTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAAC 750
     264 ..AsnIleAsnAsnLeuAspPhe.....PheSerLeuValSer 275

     751 TTGTTAAGAAAAACTATTCCGGAAGAAAAGTTA......ATTTCAATTTC 794
     276 LeuPheHisIleTyrIleProLysLysLysLysAsnTyrThrIleVa 292

     795 TGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCT 844
     292 lLysAsnAlaAsnAspIleTyrAsnSerAsnAsnSerTyrAspGlyIleI 309

     845 GTAAA.....GATGAAGAATCTCCATATAACACTAAA 876
     309 leLysAspArgGlnGluAsnIleAspGluGluGluGluLysAsnGlnGlu 325

     877 TTTTTGTCT..... 885
```

```
      |||:::
 326 HisLeuAsnLeuIleProPheLeuAspIleIleLysAsnCysIleLeuLy 342

 886 ..............................GAACAAATAGAAACAAATA 904
                                   :::::::::::::|||:
 342 sAsnIleCysIleGlnIleAsnHisLeuMetThrAsnValLysAsnAsnA 359

 905 AAGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAAT 954
     ::|||   |||:::                                |||
 359 snGlu...HisLysAsp..............................Asn 364

 955 ATTTTTAATACAGCAAAGGAGAAAATAGATCTTGTATTTATTCAAACATA 1004
     :::|||:::   |||:::|||      ::::::|||   :::   ||
 365 ProIleAsnValAsnLysLysLysLysLysValThrPheHisAspIleTy 381

1005 C............AATTTAGAAACTACAAATCCAGATATAATGGTAGATA 1042
     |           |||:::                |||     :::
 381 rGlyGluValIleAsnIleLeuTyrIleIleLysAspAspAspIleLysA 398

1043 TGTACTTATCCCATTTATATTTTGGTTTA..................... 1071
       |||:::::::::::::::      |||
 398 snTyrIleAsnTyrIleHisValIleLeuAspThrLeuTyrLysAsnIle 414

1072 ..............AAATATAAC........................... 1080
                     |||||||
 415 HisLysHisIleHisHisTyrAsnTyrThrTyrSerIleLysLeuLeuAr 431

1081 ....ATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGAT 1126
         |||:::   ||||||:::|||:::::::::|||:::
 431 gTyrIleAsnThrIleLeuSerPheAsnIleAsnHisSer.......... 444

1127 TTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACAT 1176
                                        |||:::   |||:::
 445 ......................TyrIleSerAsnGlyGlnLysIleArg 453

1177 GATAAAAATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTT 1226
     :::::::::::|||   :::|||      :::      |||   :::
 454 GluGluGluLysAsnValSerArgTyrPheSerSerLeuIleAspAsnAs 470

1227 ATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTA.......... 1266
      |||:::      :::      :::   ||||||:::
 470 nPheLeuTyrIleArgLysGluAsnPheSerPheAsnTyrValHisLysA 487

1267 ......................GATATTTTT...CTTACAAATATTTGG 1290
                            |||:::::      |||||||||
 487 snTyrAsnIleLeuTyrAsnTrpAspMetTyrAsnThrThrAsnIleAsn 503

1291 AAACATTTAAATCCTGAA.........GTACAAACTCCAAAAGACCTTAC 1331
     :::::::::|||   :::            |||      :::|||   ::
 504 AsnAsnIleAsnHisAspLysAsnIleGlnGlnTyrThrAsnAspAspAs 520

1332 TATAACTGAAAACCCTGAAGACTGTAGCACAATAGATGAATAT......G 1375
     :   |||         ||||||   :::::::::::::::|||
 520 nThrThrIleTyrIleGluAspLeuAsnAsnLeuGlnAsnTyrAsnLeuA 537

1376 TTCCAGGACTCGTTATTCCAACC.................ATAGGGATA 1407
           |||:::      |||                  |||:::::::
 537 snLysLysLeuLeuTyrLysThrAsnGluAspIleTyrAspIleSerPhe 553

1408 TATTACAAACAC.............................AATGATGC 1427
          ||||||                             |||||||::
 554 MetProLysHisIleGluHisAsnAsnAsnIleLeuAsnAspAsnAspVa 570

1428 TATATGGAAAACTAGATCTTATTCA...ATTCATGCACCTGGTGTAGACA 1474
     :   |||::::::::::|||   |||      |||   |||:::|||   :
 570 lThrTrpAsnAsnAsnSerSerSerAsnTyrHisHisProAsnValTyrL 587

1475 GATATGAATGGGAC..................................... 1488
     ::   ::::::|||
 587 ysGluGlnTyrAspLysMetGlyThrLeuProIleIleLysAsnSerPro 603

1488 ..................................................... 1488
```

```
 604 AsnGluTyrAsnLeuIleSerLysAspIleGluSerTyrLysTyrIleIl 620

1489 ................TTGGTCAAAGTATGCTATGAAAAAATATGCGATG 1522
                        |||:::   :::      :::||||||
 620 eLysPheSerIleGluLeuLeuTyrIleIleValLysLysIle....... 634

1523 GGAAAGCAGCCCATTATTATAACACTGACTATAAAGAAAGCTCTATTATT 1572
       |||   :::|||:::   ||||||   :::   ::::::::::||||||
 635 ..LysTyrValHisHisThrAsnThr...PheAlaAspGluGluIleIle 649

1573 ATA 1575
     :::
 650 Leu 650
```

```
seq_name: sp_invertebrate:O97291

seq_documentation_block:
ID   O97291      PRELIMINARY;      PRT;  1946 AA.
AC   O97291;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 231.8 KDA PROTEIN.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=99376085; PubMed=10448855;
RA   Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA   Churcher C.M., Craig A., Davies R.M, Devlin K., Feltwell T.,
RA   Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA   Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA   Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA   Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA   Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT   "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT   falciparum .";
RL   Nature 400:532-538(1999).
DR   EMBL; AL034559; CAB39037.2; -.
DR   InterPro; IPR002048; EF-hand.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW   Hypothetical protein.
SQ   SEQUENCE   1946 AA;  231792 MW;  59AC248AB9808E34 CRC64;


alignment_scores:
                Quality:  150.50            Length:    629
                  Ratio:  0.510               Gaps:     26
 Percent Similarity:  46.900   Percent Identity: 19.873

alignment_block:
 US-09-579-383-2 x O97291   ..

 Align seg 1/1  to: O97291  from: 1  to: 1946

      46 TATTCTGCAAATTCCAGAACCTTGAAAGGAAAAAATAATATAAATAATTC 95
         |||      |||:::|||         :::|||||||||   ::::::::
    1030 TyrArgAspAsnAsnArg.........AsnLysAsnAsnTyrArgAspAs 1043

      96 ATTGGGAATAATACGGGAAAATAAAAATAAAACTCATCAAACGGAAATAC 145
         :                  |||:::||||||::::::      :::   :
    1043 n..................AsnArgAsnLysAsnAsnTyrArgAspLysA 1054

     146 ATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAAT...TTTGTAGAA 192
         ::::::::   ::::::   |||:::||||||:::|||         :::
    1054 snHisAsnSerAsnAsnAsnLysAsnAsnAsnLysAsnLysAsnAsnAsn 1070

     193 TATGGATCTTATTGCGGAGATGGGTGTAACTCTAGAATTACA........ 234
         |||      |||      ::::::      |||:::||||||
    1071 TyrTyrTyrTyrGlnHisAsnAsnAsnLeuSerHisIleThrValLeuGl 1087
```

```
 235 .AAAAATAATAAAAATATAAATAAAAATGATAGAAAATCACCAAGACAAA 283
     |||:::||||||   :::||||||:::   :::         ::::::|
1087 uLysLysAsnLysAlaLeuAsnLysGlnValLysTyrLeuGluAsnLysI 1104

 284 TTTTA.....................................GAGGAG 294
     |||||                                     |||:::
1104 leLeuValGlnLysLysLysGluMetSerPheCysLysAsnLysGluLys 1120

 295 TATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAA 344
     |||||||||:::|||   ::::::|||      |||      :::   ::
1121 TyrLysLysLysLysIleSerLeuIleAsnGluTyrGluLysLysLeuAs 1137

 345 CAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGT 394
     ::::                  :::||||||   |||   :::   :
1137 pAsn.................IleIleIleAspPheAsnLysLeuLysG 1148

 395 CAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCT 444
     ::                    ::::::      :::   |||::::::
1148 luAsnCysIleThrLysGluLysLysLeuAlaLysMetGluAspIleThr 1164

 445 AGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGA 494
     :::   :::|||      :::::::|||   |||   |||      :::::
1165 LysTyrIleAsn......GluGlnPheSerLeuSerLys......IleGl 1177

 495 ATATGAAACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAG 544
     ::::|||                                     :::||||:
1177 nPheGlu.................................AsnLysM 1182

 545 TACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATG 594
     :::::      ||||||:::|||   :::      :::|||   ||||||
1182 etAsnGluTyrValIlePheLeuLysLysLysAspSerGluIleTyrMet 1198

 595 ATA..............GATATAGAAAAAGAAATTGATTATGTGGATAA 629
     :::               :::   ||||||   |||   |||   |||::
1199 LeuLysGluLeuIleLysGluLysGluLysThrIleLeuTyrAsnAspAs 1215

 630 AATATTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACT 679
     :|||||||||      :::      |||:::|||   :::
1215 nIleLeuLysGlnTyrLysLys...AspValAspAspIle..... 1227

 680 GGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAAT 729
                             |||:::|||   |||::::::
1228 .........................LeuLysGlu...AsnIleGlu... 1233

 730 TATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAA 779
           |||:::   ::::::   :::|||         :::|||::
1234 .........LysIleAspAspIleLysLysLysLeuLysThrGlnGluGl 1247

 780 GTTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAG 829
     ::::||||||        :::::::::      ::::::   :::::::
1247 uIleIleSer.........GlnLysAspArgGlnIleGluThrLeuGluA 1261

 830 GAGTTGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT 879
     ::      ::::::   |||::::::   :::   ::::::::::::::::
1261 snAsnLeuLysIleGlyLysGluLysIleAsnLysPheAspAsnGluIle 1277

 880 TTGTCTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGAT 929
        :::   |||   :::   |||   ::::::::::::   :::   ::
1278 GlnLysLeuGlnTyrLysIleAsnIleHisIleGluLysGluThrGluIl 1294

 930 GTTATCAGCAGGAACTTTTATTAATATTTTTAATACA...GCAAAGGAGA 976
     :   :::::::::   :::   :::      ||||||      :::::::|
1294 eLysLysThrAsnAspIleGluLysGluLysAsnThrLysLeuSerAsnL 1311

 977 AAATAGATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAATCCA 1026
     ||:::|||::::::               |||   |||:::::::|||
1311 ysPheAspIleIle.............AsnLysGluAsnAsnAsnLeu 1322

1027 GATATAATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATA 1076
              ::::::      |||::::::
1323 HisAsnLysIleGluThrLeuLeuLysAsnGluLysGluIleSerThrGl 1339

1077 TAACATCACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGAT 1126
```

```
           |||:::   :::|||               |||   |||:::      :
1339 uAsnValLysLeuIle.............GluThrAsnLysThrLeuT 1351

1127 TTAGTCCCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACAT 1176
     ::      ||||||::::::|||   :::   :::   :::|||:::
1351 yrIle...GluAsnGluLysLeuSerAsnAspLeuLysAsnThrLeuLys 1366

1177 GATAAAAATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTT 1226
     :::|||::::::   :::::::::   :::   |||:::      |||::
1367 GluLysGluLysIleGlnGluAsnTyrAsnLysIleAsnLysGluHisIl 1383

1227 ATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATATTTTTC 1276
     :   :::   |||:::|||   :::
1383 eLysIleIleGluGluLeuGlnSer..................... 1391

1277 TTACAAATATTTGGAAACATTTAAATCCTGAAGTACAAACTCCAAAAGAC 1326
                                 |||::::::
1392 ......................TyrLysGluIleLysGluLysHisMet 1400

1327 CTTACTATAACTGAAAACCCTGAAGACTGTAGCACAATAGATGAATATGT 1376
           ||||||:::      |||:::      :::::::|||::::::::
1401 LysGluIleThrGlnIleLysGluGlnIleHisAsnLeuAspLysHisIl 1417

1377 TCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAACACAATGATG 1426
     :      :::   |||                        |||:::|||
1417 eAlaGlnIleGlnIleGlu.................LysAsnAsn.... 1426

1427 CTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGA 1476

1426 ..................................................  1426

1477 TATGAATGGGACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAA 1526
     :::|||   :::   :::|||         ||||||      :::   ||
1427 PheGluGluSerTyrLeuLys.........GluLysAsnGluAsnGluLy 1440

1527 AGCAGCCCATTATTATAACACTGACTATAAAGAAAGCTCT.......... 1566
     |   ::::::      :::      |||||||||   |||
1440 sMetSerAsnIleLeuGluGluLysTyrLysGluLeuSerThrTyrGluI 1457

1566 ..................................................  1566

1457 leAspLysAsnIleAsnLysIleLysIleGluAspLeuGluLysAspLys 1473

1567 ......ATTATTATATGGAAAGGGGAACCATAT..............TT 1595
           |||:::::::   |||:::|||
1474 GluAsnIleLeuLeuThrLysAsnGluGluIleAsnAsnLeuLysGluGl 1490

1596 AATTAAATGGTGGCAACAAGGACCTCCGGAAGGTCAGGCACTAGAGTCAT 1645
          |||      ||||||         :::   ::::::|||   :::
1490 uTyrLysMetValGlnGlnHisLeuGluAspThrAsnValLeuTyrGluL 1507

1646 ACACAAAACTAGAT.............................GCA 1662
           ||||||
1507 ysGlnLysLeuAlaIleAspThrIleThrLysGluLysAsnAsnIleIle 1523

1663 TCCAAATGTCCAGGGATAGAAGAATGGAATAAAAAATATCCACATAAACC 1712
     :::::::|||      |||::::::   |||||||||      :::|||
1524 AsnGluCysAspLysIleLysAsnLysAsnLysLysLeuAsnAsnLysLe 1540

1713 ACTAGAAGTAGAGGAACAATATGAACAAGAAGTGGAT 1749
           |||   :::::::::||||||:::   ::::::
1540 uLysGluAsnGlnAsnAsnTyrGluHisThrLeuAsn 1552
```

```
seq_name: sp_invertebrate:O77393

seq_documentation_block:
ID   O77393      PRELIMINARY;      PRT;  2423 AA.
AC   O77393;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 293.7 KDA PROTEIN.
OS   Plasmodium falciparum (isolate 3D7).
```

```
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=99376085; PubMed=10448855;
RA   Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA   Churcher C.M., Craig A., Davies R.M, Devlin K., Feltwell T.,
RA   Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA   Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA   Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA   Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA   Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT   "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT   falciparum .";
RL   Nature 400:532-538(1999).
DR   EMBL; Z98551; CAB11149.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2423 AA;  293711 MW;  0FED1B18FC27A2F6 CRC64;


alignment_scores:
           Quality:  150.50            Length:    655
             Ratio:  0.514              Gaps:     38
 Percent Similarity:  44.733   Percent Identity: 19.847

alignment_block:
 US-09-579-383-2 x O77393   ..

 Align seg 1/1  to: O77393  from: 1  to: 2423

      76 AAAAATAATATAAATAATTCATTGGGAATA...ATACGGGAAAATAAAAA 122
         ||||||:::::::::::::        :::   ||||||:::::|||||
    1542 LysAsnLysValAspLysIleTyrGluValGluIleArgLysGluLysAs 1558

     123 TAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTTAAATCGA 172
         ||||   :::::::::::::|||::::::   :::|||:::         |
    1558 nLysLysAsnLysAsnArgIleAsnLys...TyrSerTyr.........A 1571

     173 ATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT......GGG 216
         |||||:::::::::::::                        |||
    1571 snAsnLysGluTyrIle..............LeuLysAspLeuArgArg 1582

     217 TGTAACTCTAGAATTACAAAAAATAATAAA..................... 246
         |||::::::::::|||   ||||||||||||
    1583 CysGlnLysLysIleLysLysAsnAsnLysIleIleIleLysIleAsnAs 1599

     247 ................AATATAAATAAAAATGATAGAAAATCACCAAGAC 280
                         |||:::|||:::         :::      |||:
    1599 nLysCysAsnMetTyrAsnMetAsnAsnIleTyrSerArgIleLeuArgA 1616

     281 AAATTTTAGAGGAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATAC 330
         :::::|||:::::::   ||||||||||||:::   :::::::::::|||
    1616 snPheLeuAsnLysAsnLysLysArgLysAsnLysLeuPheSerSerTyr 1632

     331 TATGGTTCATGGAACAGTCAAGGTGATAGAGCA......AAACATATGAT 374
            :::      |||:::         |||:::      |||||||:::||
    1633 ValAsnIleLeuAsnAspCysLeuHisArgThrIleArgLysHisIleIl 1649

     375 TGATTCAAACCCAATGGTGTCAATTTTATATATTGCATTTGCTCGCATTA 424
         |   :::|||                  :::|||
    1649 eLysAsnAsn..................PheIle.............. 1654

     425 ATATGTTATATGATGTATCTAGACCATTTAATGGAAGACAAAGATTCCTA 474
                               :::|||:::::::|||:::
    1655 .........................TyrAsnAsnLysGlnAsn...... 1660

     475 TTAAGAAAACACGGCTTAGAATATGAAACCTATGGTATGATGCTTAATGA 524

    1660 ................................................. 1660

     525 AATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCT 574
                      :::|||:::      |||:::      |||   ||||

    1661 .............AsnLysIleValGlnAspIleLysGluLeuTyrSerL 1673

     575 TAGGTGGAGAAACCTATATGATAGATATAGAA..................... 606
         ||:::::::         :::   |||||||||
    1673 euSerSerValHisAsnIleArgAspIleGluCysGlyTyrGlnGluArg 1689

     607 ......AAAGAAATTGATTATGTGGATAAAATATTG............AA 638
               |||||||:::          |||   |||:::            ||
    1690 MetArgLysGluMetLeuLysPheAspIleIleIleAspHisLysGluLy 1706

     639 GCTTGTTAATGATTTTGAT.........TTAGATGGTGTAGATATTGAC. 678
         ||||:::|||:::|||:::          :::|||:::::::::::::
    1706 sLeuIleAsnSerPheAsnPheValCysValAspAsnIleSerLeuGlyL 1723

     678 ................................................. 678

    1723 euHisAsnAsnAsnAsnAsnAsnLysAlaLeuGluAsnLysGlyLysSer 1739

     679 ......TGGGAACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTT 722
               :::::::|||:::|||:::       |||:::|||
    1740 GlyAlaTyrLysProTyrGlySerAsnLysAsnIleAsn........... 1752

     723 TTCAAATTATTATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGG 772
                         |||   ||||||   ::::::|||:::|||
    1753 ................LysAsnIleAsn...IleAsnLysAsnIleAsnI 1763

     773 AAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGC 822
           :::|||   |||:::|||:::   :::   ||| ::::::   :::::
    1763 leAsnLysAsnIleAsnIleAsnLysAsnIleAsn.IleAsnLysAsnVa 1779

     823 GTTTCAGGAGTTGCATCTTTCTGTAAAGATG.........AAGAATCTCC 863
         :      :::   ::::::::::::::::::          |||||||:::
    1779 lAsnIleAsnLysAsnValAsnIleAsnGlnAsnIleAsnLysAsnIleA 1796

     864 ATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTAC 913
         ::|||:::::::|||:::   ::::::::|||   ::::::::::::|||
    1796 snIleAsnIleAsnIleAsnIleAspLysGlySerAsnMetAsnAsnPro 1812

     914 ATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTT.............. 948
           ::::::|||:::      ::::::::::::
    1813 CysAsnHisGlnAsnLysAsnArgGluAspIleHisTyrAsnAsnHisTy 1829

     949 ..........ATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCTTG 988
                    |||   :::  |||      ::::::|||   :::|||:
    1829 rLysHisMetIleTyrPheHisAsnLysHisAsnAspLysSerGluLeuL 1846

     989 TA.............................................. 990
         ::
    1846 euLysLysAsnSerGlyGlyLeuGluTyrAsnHisGluGluArgLysTyr 1862

     990 ................................................. 990

    1863 LysGluLysIleTyrSerCysLysArgAsnPheGluIleGluAsnIleAs 1879

     991 ....TTTATTCAA.....................ACATACAATTTAGAAA 1015
             ||||||                        |||:::   :::
    1879 nTyrPheIleTyrArgAsnIleTyrArgProIleLysTyrHisTyrAspI 1896

    1016 CTACAAATCCAGATATAATGGTAGATATGTACTTATCCCAT......... 1056
           ||||||      |||:::::::::   :::|||   :::
    1896 leThrAsn...LysIleLeuIleAsnSerPheLeuIleAsnMetAsnLeu 1911

    1057 ...TTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTTTTC 1103
            :::||||||         |||   :::      ||||||      ::
    1912 LeuIleTyrPhe.........TyrLeuLeuLysCysIleLeuPheAspAs 1925

    1104 ATTAGAACATAACAGAGGTGGA..............TTTAGTCCCGAAA 1138
         :::::::   |||   :::                  |||:::   :::|
    1925 pMetAspAspAsnTyrAsnPheLeuCysAsnIleLeuPheAsnLysArgA 1942

    1139 ATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAA 1188
         |||||            ::::::::::|||   |||::::::   :::
    1942 snLys............IleIleAsnLys...IleGluAsnAspAspTyr 1953
```

```
    1189 AATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTTATGAAAGA 1238
         |||:::|||   :::|||               |||:::|||:::::::
    1954 AsnAspAsnCysCysAsp.............HisIlePheValAsnLy 1965

    1239 ACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACA...AATA 1285
         :           ::::::      |||:::   ||||||:::   |||:
    1965 s............AsnLysHisMetValGlnTyrPheLeuAsnLysAsnP 1978

    1286 TTTGGAAACATTTAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATA 1335
         :::::        |||                 |||:::      :::
    1978 heTyrThrSerCysAsn.............IleLysAsnLysCysVal 1989

    1336 ACTGAAAACCCT.........GAAGACTGTAGCACAATAGATGAATATGT 1376
            :::|||              |||   |||||||||
    1990 AspLysAsnLysLeuTyrTyrThrAspIleSerThrIle.......... 2002

    1377 TCCAGGACTCGTTATTCCAACCATAGGGATATATTACAAACACAATGATG 1426
                                  ::::::|||   |||
    2003 ........................PhePheTyrIleHis....... 2007

    1427 CTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGA 1476
                             ::::::|||::::::|||
    2008 ...................PheLysIleAsnValPro............ 2013

    1477 TATGAATGGGACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAA 1526
                  |||:::::::::   :::            |||   ||
    2014 .........LysLeuLeuArgIlePhePhe............AspHisLy 2023

    1527 AGCAGCCCATTATTATAACACTGACTATAAAGAAAGCTCTATTATT.... 1572
         |:::::::::|||||||||      |||:::   :::|||::::::
    2023 sValValAsnTyrTyrAsnAspIleTyrArgLeuThrSerLeuLeuTyrP 2040

    1573 ........................................ATATGG 1578
                                                  :::
    2040 heSerLeuLeuAsnLeuAsnTyrIlePheLeuIlePheSerThrLeuCys 2056

    1579 AAAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCTCCG..... 1623
         :::         :::      |||   |||:::   |||   |||
    2057 ArgProThrIlePheGlnCysLysLysTrpLysGlyGlyAsnProSerTh 2073

    1624 ....GAAGGTCAG 1632
             |||||||||
    2073 rAlaGluGlyGln 2077
```

seq_name: sp_fungi:Q06350

seq_documentation_block:

```
ID   Q06350      PRELIMINARY;      PRT;   511 AA.
AC   Q06350;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILAR TO CHITINASE.
GN   YDR371W OR D9481.7.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C (AB972);
RA   Ding H.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C (AB972);
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA   Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA   Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA   Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA   Taich A., Trevaskis E., Vignati D., Wilcox L., Wohldman P., Vaudin M.,
RA   Wilson R., Waterston R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C (AB972);
RA   Waterston R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C (AB972);
RA   Jia Y., Cherry J.M.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; U28373; AAB64807.1; -.
DR   SGD; S0002779; YDR371W.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   Pfam; PF00704; Glyco_hydro_18; 2.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Glycosidase; Hydrolase.
SQ   SEQUENCE   511 AA;  59118 MW;  238FF79963BF0D63 CRC64;
```

alignment_scores:

```
            Quality:  149.00            Length:    599
              Ratio:  0.598               Gaps:     28
 Percent Similarity:  41.569   Percent Identity: 18.698
```

alignment_block:

US-09-579-383-2 x Q06350 ..

Align seg 1/1 to: Q06350 from: 1 to: 511

```
    140 AAATACATGAGTCTTTTTCACATCTTAAATCGAATAATAGTAATTTTGTA 189
        ::::::   |||         :::::::::::::::::::::::::|||
      7 GlnHisArgSerLysSerSerLeuValSerHisLeuLeuIleLeuLeuIl 23

    190 GAATATGGATCTTATTGCGGAGATGGGTGT...AACTCTAGAATTACAAA 236
            :::   :::|||   ||||||  |||      :::::|||   ||
     23 ePheIleThrIleIleIleGluMet..CysLeuTyrAsnLysIlePheLy 39

    237 AAATAATAAA.................AATATAAATAAAAATGATAGAA 268
        ||||::::::                 |||:::|||::::::   :::::
     39 sAsnGlnArgSerAspAspIleArgAspAsnPheAsnAsnGlyGlyHisA 56

    269 AATCACCAAGACAAATTTTAGAGGAGTATAAAAAA...AGGAAACAAGGT 315
        ::   |||   ::::::   ::::::           |||   ::::::
     56 rgValProSerAsnValGlnAsnHisGlyThrHisIleArgAspGluAla 72

    316 ATTATAGCAGGA...TACTATGGTTCATGGAACAGTCAAGGTGATAGAGC 362
        :::|||:::|||   ||||||:::::::|||:::            |||
     73 PheIleSerGlyValTyrTyrSerAsnTrpSerProTyrLysProArgPh 89

    363 AAAACATATGATTGATTCAAAC...CCAATGGTGTCAATTTTATATATTG 409
                    |||   |||      :::||||||   :::|||   |
     89 eHisPheProHisAspIleAsnLeuLysGlnValSerHisIleTyrTyrA 106

    410 CATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCATTTAATGGA 459
        |||||   :::||||||
    106 laPhePheLysIleAsnSer............................. 112

    460 AGACAAAGATTCCTATTAAGAAAACACGGCTTAGAA.............. 495
                            |||      |||:::|||
    113 ..................ArgThrGlyGlyIleGluAsnThrAspSerTr 123

    496 ......................TATGAAACCTATGGTATGATGCTTAATG 523
                               |||::::::   ::::::      :::|
    123 pSerAspLeuGluMetAsnLeuTyrLysSerLeuAlaIleLysAsnSerG 140

    524 AAATT.................AGACGTATCAGAAAAGTACGTCCA... 552
        ||:::                  :::   ::::::::::::   |||
    140 luLeuIleLysGluSerSerAsnAsnSerValGlnAsnIleLeuProLeu 156
```

```
 553 ..............................................GA 554

 157 GlyCysIleGlyGluLeuPheTyrLeuLysAsnThrCysSerAspLysLy 173

 555 TGTAATTATTCTTTTATCCTTAGGTGGA............GAAACCTATA 592
          :::::::::|||:::||||||            |||::::::.
 173 sPheLysValIleMetSerIleGlyGlyTrpSerAspSerGluAsnPheL 190

 593 TGATAGATATAGAAAAAGAA.........ATTGATTATGTGGATAAAATA 633
       |||   |||:::   :::          ::::::||||||:::
 190 ysIleIleIleLysAspAspLysLeuLeuGlnAsnPheValAspSerSer 206

 634 TTGAAGCTTGTTAATGATTTTGATTTAGATGGTGTAGATATTGACTGGGA 683
     ::::::   :::      :::   :::||||||:::|||:::||||||||
 207 ValGluThrMetPheArgLeuGlyPheAspGlyIleAspLeuAspTrpGl 223

 684 ACCACATGGGAAGTTTTACAACTTAAATGAATTAAATTTTTCAAATTATT 733
     |      |||               |||:::   :::      :::   |
 223 uPheProGly..............AsnAsnGluSerGluProArgGlyT 235

 734 ATATTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTA 783
     ||:::||||||::::::::::||||||                ||||||
 235 yrLeuLysLeuValArgMetLeuArgLeu..............LysLeu 246

 784 ATTTCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGT 833
        |||:::
 247 AsnSerLeu..................................... 249

 834 TGCATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTGT 883
                                    ::::::::::::::::
 250 ................................GluSerGlnIlePheG 255

 884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
     ::::::::   |||   :::   :::|||      ||||||      :::
 255 lyLysArgThrGluAspHisPheGlnLeuSerIleAlaAlaProAlaPhe 271

 934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
     :::        |||        ::::::          :::   :::||
 272 LysAspLysLeuPheTyrLeuProIleThrGluIleAspGlnTyrValAs 288

 984 TCTTGTATTTATTCAAACATACAATTTA..............GAAACTA 1018
     |         ::::::|||||||:::                 |||||||
 288 pTyrTrpAsnMetMetThrTyrAspTyrTyrGlySerTrpSerGluThrT 305

1019 CAAATCCAGATATAATGGTAGATATGTACTTATCCCATTTATAT...... 1062
     ||:::                    |||   |||:::|||:::
 305 hrGly.....................TyrHisSerAsnLeuPheSerGlu 314

1063 ..................TTTGGTTTAAAATATAACATC........... 1083
                       |||::::::   |||   |||
 315 ThrGluLeuAsnGlyAsnPheAlaMetHisTyrMetIleAspArgPheGl 331

1084 .............ACAATCATATTAGGT.....................T 1099
                     ::::::|||||                      |
 331 yValAsnSerArgLysLeuValLeuGlyMetAlaAlaTyrGlyArgSerP 348

1100 TTTCATTAGAACATAACAGAGGTGGATTTAGTCCC......GAAAATAAA 1143
     ||   ::::::   |||:::      |||:::|||      :::|||
 348 heHisIleLysAspAsnLys......PheGluProPheAsnGlnAsnThr 362

1144 GAATTATTAGAATTGGTAGGAAAAACAATACATGATAAAAATCAAAATAA 1193
        |||::::::   :::   |||   :::          :::::::::
 363 ValLeuIleAsnLysIlePheLysGlyValGlyLysProThrLysGluIl 379

1194 TAATAGGGCAGATGGT...ATAGGGATATGGCATTTATTTATGAAAGAAC 1240
      ::::::||||||||||      |||||||||
 379 eAspLysAlaAspGlyLysGluGlyIleTrp.................. 389

1241 AATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAAATATTTGG 1290

 389 ..............................................  389

1291 AAACATTTAAATCCTGAAGTACAAACTCCAAAAGACCTTACTATAACTGA 1340
                  |||      :::   ||||||
  390 ............ProTyrLysAsnLeuProLys.............. 396

1341 AAACCCTGAAGACTGTAGCACAATAGATGAATATGTTCCAGGACTCGTTA 1390
                       :::||||||::::::|||   |||      |||
  397 .............IleGlyThrIleGluGlnTyrAspProLysTyrValS 409

1391 TTCCA.....................ACCATAGGGATATATTACAAACAC 1419
                                  :::|||   |||   |||   :::
  409 erAlaTyrCysPheAspGluLysAsnSerIlePheIleSerTyrAspAsn 425

1420 AATGATGCTATATGGAAAACTAGATCTTATTCAATTCATGCACCTGGTGT 1469
       :::   ::::::              :::|||      |||
  426 ThrLysSerValLysThrLysAlaGluTyrValThrHisAsnAsnLeuGl 442

1470 AGACAGATATGAATGGGACTTGGTCAAAGTATGCTATGAAAAA....... 1512
           :::   |||:::            :::|||   :::
  442 yGlyGlyPheTrpTrpGluSerCysGlyGluAlaTyrAlaAsnGluSerA 459

1512 ..............................................  1512

  459 rgSerLeuIleAsnAlaPheAsnGluGlyLeuHisPheAsnValSerSer 475

1513 ............ATATGCGATGGGAAAGCAGCCCATTATTATAACACTGA 1550
                    :::   |||   ::::::     ||||||   :::
  476 LysProSerIlePheGlnAspValArgValLysLysTyrTyrLeuAsnLy 492

1551 CTATAAAGAAAGCTCTATTATTATATGGAAAGGGGAACCATATTTA 1596
       |||   ::::::::::::::::            :::|||||||||
  492 sTyrGlyAspGlyGlyPheLeu............SerProTyrLeu 503

seq_name: sp_bacteria:O54328

seq_documentation_block:
ID   O54328      PRELIMINARY;      PRT;   563 AA.
AC   O54328;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CHITINASE (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-BETA-POLY-N-
DE   ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).
GN   CHIA.
OS   Enterobacter sp.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Enterobacter.
OX   NCBI_TaxID=42895;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G-1;
RA   Park J.K., Okamoto T., Yamasaki Y., Tanaka K., Nakagawa T.,
RA   Kawamukai M., Matsuda H.;
RL   J. Ferment. Bioeng. 84:493-501(1997).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-
CC       ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; U35121; AAB97779.1; -.
DR   HSSP; P07254; 1CTN.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR002173; PfkB.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
KW   Glycosidase; Hydrolase.
SQ   SEQUENCE   563 AA;  61017 MW;  E455EC15B050AE37 CRC64;


alignment_scores:
            Quality:  148.50            Length:    425
```

```
            Ratio:  0.746              Gaps:     20
 Percent Similarity:  46.824   Percent Identity: 20.706

alignment_block:
 US-09-579-383-2 x O54328   ..

 Align seg 1/1  to: O54328  from: 1  to: 563

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||  ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||                 ::::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   ::::::::::|||      |||:::|||
      73 aGlyThrThrAlaLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ............................AATATAAATAAAAATGAT 264
                                     ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA........................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 ........................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuAlaP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
              |||:::|||      :::||||||:::   |||      ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
         :::         :::::::::::|||:::   :::|||:::      ||
     155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         ::      :::|||:::   |||      ::::::   :
     171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
         :::::   ::::::   |||               |||   :::   |||
     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

     439 ............................GTATCTAGACCATTTAA 455
                                    :::::::|||      ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC...........GGCTTAG 493
         :||||||:::   |||   :::      |||           :::|||:
     219 nGlyArgGluAspPheLysValSerIleHisAspProPheAlaAlaLeuG 236

     494 AATATGAAACCTATGGTATGATG....................... 516
         ::          |||:::
     236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::||||||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaArgProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
         |   |||:::||||||   |||                   :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
            :::      :::|||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

     693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
         ||||            ||||||::::::   :::   ::::::   |||:
     319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

     737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
         ::   |||::::::   ||||||      :::   :::
     336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............ 348

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
                                      |||||||::   :::|||
     349 ..........................LeuSerAlaGluThrGly..... 354

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTT...TTGT 883
                                                 |||:::   |||:
     355 ................................ArgLysTyrGluLeuT 360

     884 CTGAACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTA 933
         :::::   |||::::::::::|||:::          :::|||
     360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla...... 374

     934 TCAGCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGA 983
                                    :::|||:::|||:::::::::::||
     375 ........................TyrAsnValAlaGlnAsnSerMetAs 383

     984 TCTTGTATTTATTCAAACATACAAT 1008
         |   :::|||:::::::::|||:::
     383 pHisIlePheLeuMetSerTyrAsp 391

seq_name: sp_invertebrate:Q25802

seq_documentation_block:
ID   Q25802      PRELIMINARY;      PRT;   960 AA.
AC   Q25802;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FRAMESHIFT.
GN   RPOD.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C10;
RX   MEDLINE=96346169; PubMed=8757284;
RA   Wilson R.J.M., Denny P.W., Preiser P.R., Rangachari K., Roberts K.,
RA   Roy A., Whyte A., Strath M., Moore D.J., Moore P.W., Williamson D.H.;
RT   "Complete gene map of the plastid-like DNA of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 261:155-172(1996).
DR   EMBL; X95275; CAA64574.1; -.
DR   InterPro; IPR000722; RNA_pol_A.
DR   Pfam; PF00623; RNA_pol_A; 1.
SQ   SEQUENCE   960 AA;  117987 MW;  F41782D73607F35D CRC64;


alignment_scores:
            Quality:  148.50              Length:    661
              Ratio:  0.488                 Gaps:     43
 Percent Similarity:  45.991   Percent Identity: 22.995

alignment_block:
```

US-09-579-383-2/rev x Q25802 ..

Align seg 1/1 to: Q25802 from: 1 to: 960

```
1561 TTTCTTTATAGTCAGTGTTATAATAATGGGCTGCTTTCCCATCGC..... 1517
     |||   :::   :::   ||||||      :::|||:::   :::
   3 PheTyrPhePheAsnLysTyrAsnLeuLysIleLeuGluLysLysLeuLe 19

1516 .ATATTTTTTCATAGCATACTTTGACCAAGTCCCATTCATATCTGTCTAC 1468
      |||:::|||                              :::|||   :::
  19 uIleIlePhe.....................LysTyrAsnIleSerPheL 29

1467 ACCAGGTGCATGAATTGAATAAGATCTAGTTTTCCATATAGCATCATTGT 1418
          :::|||||||||| :::      ::::::   ||||||:::
  29 ysIleLeuHisGluLeu.LeuTyrLeuGlyTyrGluTyrSerPhe..... 43

1417 GTTTGTAATATATCCCTATGGTTGGAATAACGAGTCCTGGAACATATTCA 1368
           :::|||   |||:::   |||   ::::::::::::   ||||
  44 .......LeuTyrAsnTyrSerLeuAsnIleLysAspPheSerAsnPheI 58

1367 TCTATTGTGCTACAGTCTTCAGGGTTTTCAGTTATAGTAAGGTCTTTTGG 1318
     |||||         :::::::::      :::   ::::::::::::
  58 leTyrLeuLeuIleLeuTyrLys...AsnLysIleAsnAsnIleTyrAsn 73

1317 AGTTTGTACTTCAGGATTTAAATGTTTCCAAATATTTGTAAGAAAAATAT 1268
     :::   |||::::::|||                          ||||||
  74 AsnLysTyrTyrGluIle.......................LysAsnAs 82

1267 CTACATCGAATGATCCAGTTGGTAATTGTTCTTTC.........ATAAAT 1227
      |||||| |||     ::::::|||      :::         ||||||
  82 nTyrIle.AsnValPheLeuAsnAsnTyrTyrTyrLeuLysValIleAsn 98

1226 AAATGCCATATCCCTATACCATCTGCCCTATTATTATTTTGATTTTTATC 1177
     |||           |||   :::   |||
  99 Lys............IleGlnGlyIleLeu..................... 104

1176 ATGTATTGTTTTTCCTACCAATTCTAATAATTCTTTATTTTCGGGACTAA 1127
                              ||||||:::|||::::::   :::|
 105 .........................AsnAsnAsnLeuTyrAsnLysIleA 113

1126 ATCCACCTCTGTTATGTTCTAATGAAAAACCTAATATGATTGTGATGTTA 1077
     |||||      :::   ||||||                  :::::::|||
 113 snPro......IleTyrSerAsn.................LeuPheLeu 121

1076 TATTTTAAACCAAAATATAAA........................... 1056
     :::|||:::   |||   |||
 122 PhePheAsnAsnLysIleLysIleLysTyrSerGlnLeuGlnGlnLeuIl 138

1055 ....TGGGATAAGTACATATCTACCATTATATCTGGATTTGTAGTTTCTA 1010
         :::      |||||||||:::   |||:::|||::::::   :::|
 138 eGlyTyrLysGlyTyrIleSerAsn...IleLysGlyMetIleTyrGluL 154

1009 AATTGTATGTT......TGAATAAATACAAGATCTATTTTCTCCTTTGCT 966
     ||      :::         ||||||      :::|||::::::::::
 154 ysProValIleAsnAsnTyrIleAsnGluLeuAsnIleTyrGluTyrIle 170

 966 ..................................................... 966

 171 LeuSerCysTyrGlySerLysLysGlyIleIleAspThrAlaLeuLysTh 187

 965 ................GTATTAAAAATATTAATAAAAGTTCCTGCTGATA 932
                     :::   |||   ||||||::::::   ::::::
 187 rAlaAspSerGlyTyrLeuThrLysArgLeuIleAsnIleThrSerAsnP 204

 931 ACATCGCTGCTGCCCTA...TGTAATTCTTTATTTGTTTCTATTTGTTCA 885
       |||         |||   |||:::|||   |||:::   :::
 204 heIleIleLysGluLeuAsnCysLysSerProPheIle...LeuLysTyr 219

 884 GACAAAAATTTAGTGTTATATGGAGAT..................... 858
           |||:::   :::||||||:::
 220 IleLeuAsnMetAspIleTyrGlyAsnIleIleLeuProLeuAsnIleLe 236

 857 ....TCTTCATCTTTACAGAAAGATGCAACTCCTGAAACGCATGATAATG 812
```

```
         :::   ||||||::::::
236 uArgPheLysIleLeuGlnAsnAsnIle..................... 245

811 CAGCATTTGATGAACCAGAAATTGAAATTAACTTTTCTTCCGGAATA... 765
                              :::|||:::::::::|||
246 .........................LeuAsnLeuAsnAsnGlyThrPhe 253

764 ..............................GTTTTTCTTAACAAGTTAAT 745
                                   :::||||||||||||::
254 IleTyrThrLysAsnThrTyrIleThrLysTyrIleLeuAsnLysLeuLe 270

744 TAATTTAATATAAT.................AATTTGAAA......... 722
    :|||||| ||||||                 |||:::|||
270 uAsnLeu.TyrAsnArgArgAsnIleTyrLeuAsnIleLysSerValTyr 286

721 ......AATTTAATTCATTTAAGTTGTAAAACT................. 695
         |||:::   :::      |||:::|||
287 LeuCysAsnIleTyrAsnAsnIleCysAsnThrCysLeuAsnTyrLysGl 303

694 ......................TCCCATGTGGT..........TCCCAGT 677
                           :::|||:::            |||::::
303 nLeuTyrLysTyrAsnLeuGlyGlnHisIleGlyValIleSerSerGluA 320

676 CAATATCTACACCATCTAAATCAAAATCATTAACAAGCTTCAATATTTTA 627
    ::|||||   ||||||   :::      |||   :::|||:::
320 laIleSerGluProSerThrGlnMetValLeuArgThrPheHis...... 334

626 TCCACATAATCAATTTCTTTTTCTATATCTATCATATAGGTTTCTCCACC 577
                        :::   ||||||:::
335 .....................AlaSerSerIleLeu.............. 339

576 TAAGGATAAAAGAATAATTACATCTGGACGTACTTTTCTGATACGTCTAA 527
     |||||||||      :::::::::         :::   |||   |||
340 .LysAspLysPheAsnPheAsnLysTyrLeuIleTyrLysIleTyrLeuT 356

526 TTTCATTAAGCATCATACCA............TAGGTTTCATATTCTAAG 489
      :::|||:::|||                  :::::::::::|||
356 yrLysLeuAsnIleAsnLysIlePheLysLeuIleIleAsnPheLysLys 372

488 CCGTGTTTTCTTAATAGGAATCTTTGTCTTCCATTAAATGGTCTAGATAC 439
          ::::::   ||||||   :::   :::|||   :::
373 TyrIleAsnIleLysPheAsnLeuIlePheLeuMetAsnLysIleLeuTy 389

438 ATCATAT...AACATATTAATGCGAG..............CAAATGCAA 407
     :::|||   ||||||||||::: |||       '        |||   |||
389 rAsnTyrAsnAsnIleLeuPhe.GluTyrLysTyrIleLeuGlnAsnGln 405

406 TATATAAAATTGACACCATTGGGTTTGAATCAATCATATGTTTTGCTCTA 357
    |||||||||   :::   :::      :::   |||      :::   ||
406 TyrIleLysCysAsnPheIleTyrAsnSerIleSerLysAsnPheLysTy 422

356 TCACCTTGACTGTTCCATGAACCATAGTATCCTGCTATAATACCTT.... 311
    |:::|||               ::::::|||::::::   ||||||
422 rAsnLeu..............AsnAsnIleIleIleLysTyrLeuAsnA 434

310 .....GTTTCCTTTTTT...TATACTCCTCTA.................. 287
         ::::::::::::   |||:::   :::
434 snValIleLysTyrTyrAsnTyrSerAsnIleGlnLeuLeuIleLysAsn 450

286 .........AAATTTGTCTTGGTGATTTTCTATCATTTTTATTTATATTT 246
            |||::::::|||      :::|||   :::|||||||||::
451 IleHisAsnLysTrpIleLeuTyrAsnIleTyrThrTyrTyrLeuTyrTy 467

245 TTATTATTTTTTGTAATTCTAGAGTTACACC................... 215
    :|||::::::      :::      |||:::
467 rTyrHisIleLysPheTyrAsnLeuTyrAsnLysGlyIleIleLeuAsnA 484

214 ............................CATCTCCGCAATAAGATCCAT 194
                                   |||   |||   ::::::
484 snAsnAsnAsnLysTyrAsnValIleTyrPheLeuIleAsnTyrPheAsn 500

193 ATTCTACAAAATTACTATTATTCGATTTAAGAT..............GT 159
    :::::::::|||||||||||||:::|||   :::              ::
```

```
     501 LeuPheSerAsnTyrTyrTyrLysIleTyrAsnAsnAsnTyrAsnPheIl 517

     158 GAAAAAGACTCATGTATTTCCGTTTGATGAGTTTTATTTTTA......TT 115
         :::::::::::: |||||||:::   |||:::|||||| |||      ||
     517 eAsnSerAsnTyr.TyrPheLysLysMetAsnPheIle.LeuLysAsnPh 533

     114 TTCCCGTATT...ATTCCCAATGAATTATTTATATTA......TTTTTTC 74
         |:::::|||   |||   |||:::||||||   :::      :::|||
     533 eAsnAsnIleGlnIleLeuAsnLysLeuPheTyrValAsnAsnIlePheI 550

      73 CTTTCAAGGTTCTGGAATTTGCAGAATACAAGATGGATACTATAATTAAA 24
           :::         |||         :::         :::||||||:::
     550 leTyrTyrLysTyrGluLysLysLeuPheIleTyrLeuAsnIleIleAsn 566

      23 AATATTGATATTTTAAAATTC 3
         ||||||   |||   |||:::
     567 AsnIleIleIleLysLysTyr 573

seq_name: sp_archea:Q9UWR7

seq_documentation_block:
ID   Q9UWR7      PRELIMINARY;      PRT;  1215 AA.
AC   Q9UWR7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CHITINASE.
GN   PK-CHIA.
OS   Pyrococcus kodakaraensis.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX   NCBI_TaxID=69014;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20049967; PubMed=10583986;
RA   Tanaka T., Fujiwara S., Nishikori S., Fukui T., Takagi M., Imanaka T.;
RT   "A unique chitinase with dual active sites and triple substrate
RT   binding sites from the hyperthermophilic archaeon Pyrococcus
RT   kodakaraensis KOD1.";
RL   Appl. Environ. Microbiol. 65:5338-5344(1999).
DR   EMBL; AB024740; BAA88380.1; -.
DR   HSSP; P07103; 1AIW.
DR   InterPro; IPR001919; CBD_2.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR003610; Chitin_bind3.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   Pfam; PF00553; CBD_2; 2.
DR   Pfam; PF00704; Glyco_hydro_18; 2.
DR   SMART; SM00495; ChtBD3; 1.
DR   PROSITE; PS01095; CHITINASE_18; UNKNOWN_1.
SQ   SEQUENCE   1215 AA;  134266 MW;  7A3997EEBF7A86C0 CRC64;


alignment_scores:
            Quality:  148.00            Length:    647
              Ratio:  0.576              Gaps:     28
 Percent Similarity:  39.722   Percent Identity: 17.311

alignment_block:
 US-09-579-383-2 x Q9UWR7   ..

 Align seg 1/1  to: Q9UWR7  from: 1  to: 1215

     160 CATCTTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGG 209
         |||||||||:::|||:::::   :::                   ::
     139 HisLeuLysAlaAsnThrThrTyrTyrTyrGlyValValProValLeuAl 155

     210 AGATGGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAAA 259
         :||||||
     155 aAspGly..................................... 157

     260 ATGATAGAAAATCACCAAGACAAATT.................TTAGAG 291
           :::|||   ||||||   ::::::                |||||
     158 ..SerArgGlySerProSerAsnValLeuAlaIleThrThrProLeuGlu 173

     292 GAGTATAAAAAAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATG 341
           |||:::                   :::      ||||||   |||||
     174 ProTyrArg.................ValIleValTyrTyrIleSerTr 184

     342 GAACAGTCAAGGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGG 391
         |:::      :::         |||      :::::::::   |||
     184 pGlyArgTyrAla......ArgLysPheTyrValSerAspIleProTrpG 199

     392 TGTCAATTTTATATATTGCATTTGCTCGCATTAATATGTTATATGATGTA 441
           ::::::   ::::::   :::|||   :::::::::      |||
     199 luLysValThrHisValAsnTyrAlaPheLeuAspLeuLysGluAspGly 215

     442 TCTAGACCATTTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTT 491
         :::      |||
     216 ThrValAlaPhe..................................... 219

     492 AGAATATGAAACCTATGGTATGATGCTTAAT......GAAATTAGACGTA 535
           |||:::||||||:::      ||||||         :::::::::
     220 ....TyrAspThrTyrAlaAspProLeuAsnLeuGluAlaMetLysGluT 235

     536 TCAGAAAAGTACGTCCAGATGTAATTATTCTTTTATCCTTAGGTGGAGAA 585
           ::::::      |||   |||   :::|||:::|||:::||||||
     235 yrLysArgLysTyrProAlaValLysValLeuIleSerValGlyGlyTrp 251

     586 ACC.....................TATATGATAGATATAGAAAAAGAAAT 614
         |||                              |||      |||:::
     252 ThrLeuSerLysTyrPheSerValValAlaAlaAspProAlaLysArgGl 268

     615 TGATTATGTGGATAAAATATTGAAGCTTGTTAATGATTTTGATTTAGATG 664
           :::::::::      ::::::::::::::   ::::::|||||||
     268 nArgPheAlaGluThrAlaIleGluIleLeuArgLysTyrAsnLeuAspG 285

     665 GTGTAGATATTGACTGGGAA.........CCACATGGGAAGTTTTACAAC 705
         ||:::|||||||||||||||                |||         |||
     285 lyIleAspIleAspTrpGluTyrProGlyGlyGlyGlyMetAlaGlyAsn 301

     706 TTAAATGAATTAAATTTTTCAAATTATTATATTAAATTAATTAACTTGTT 755
           ::::::   :::   ::::::   ::::::   |||::::::   ||
     302 TyrGluSerProAspAspGlyLysAsnPheValLeuLeuLeuLysAspLe 318

     756 AAGAAAAACTATTCCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAA 805
         ||||::::::::   :::
     318 uArgGluAlaLeuAspLysAla........................... 325

     806 ATGCTGCATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAA 855
                                                :::|||::::::
     326 .....................................AlaLysGluAsp 329

     856 GAATCTCCATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAA 905
         ::::::   |||
     330 HisLysAspTyr..................................... 333

     906 AGAATTACATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAATA 955
                                 :::|||:::|||:::|||      :::
     334 ......................LeuLeuThrAlaAlaThrProAlaAspP 343

     956 TTTTTAATACA.....................GCAAAGGAGAAAATAGAT 984
           ::::::                        |||::::::   :::|||
     343 roValLysAlaGlyArgIleAspTrpValGluAlaSerLysTyrLeuAsp 359

     985 CTTGTATTTATTCAAACATAC........................... 1005
           :::   |||:::||||||
     360 SerIleAsnIleMetThrTyrAspTyrHisGlyAlaTrpGluThrIleTh 376

    1006 ...............................AATTTAGAAACTACAAATC 1024
                                        |||         |||:::
     376 rGlyHisLeuAlaProLeuTyrCysAspProAsnAlaProTyrThrAspG 393

    1025 CAGATATA.....................ATGGTAGATATGTACTTA 1050
           ::::::                    |||:::   |||:::
     393 luAsnValLysTyrHisPheCysValAsnTyrThrValGlnTrpTyrIle 409
```

```
 1051 TCCCATTTATATTTTGGTTTAAAATATAACATCACAATCATATTAGGTTT 1100
      :::|||:::           |||   :::||||||:::   |||   ||
  410 GlnHisVal......ProAspLysThrLysIleThrValGlyLeuProPh 424

 1101 TTCATTAGAACATAACAGAGGTGGATTTAGTCCCGAAAATAAAGAATTAT 1150
      |     :::::::::   ::::::     |||||||||:::   |||
  424 e...TyrSerArgSerPheAlaAsnValProProGluAsnAsnGlyLeuT 440

 1151 TAGAATTGGTAGGAAAAACA..............................  1170
        :::     :::   |||
  440 yrGlnProPheSerGlyThrProAlaGlyThrTrpGlyProAlaTyrGlu 456

 1171 ..........................ATACATGATAAAAATCAAAATAA 1193
                                 :::   :::||||||||||:::::
  457 ThrTyrGlyValMetAspTyrTrpAspValAlaGluLysAsnGlnSerSe 473

 1194 T.............................................. 1194
      :
  473 rGluTyrGluTyrHisTrpAspProIleAlaGlnValAlaTrpLeuTyrS 490

 1194 .............................................. 1194

  490 erProSerLysArgIlePheIleThrPheAspAspProArgAlaIleGly 506

 1195 ......................AATAGGGCAGATGGTATAGGGATATG 1220
                            |||         |||:::   |||||
  507 IleLysValAspTyrMetLeuLysAsnGlyLeuGlyGlyValMetIleTr 523

 1221 GCATTTATTTATGAAAGAACAATTACCAACTGGATCATTCGATGTAGATA 1270
      |:::::         ::::::      |||:::::      :::
  523 pGluIleThrAlaAspArgLysProGlyThrAsnAspHisProLeu.... 538

 1271 TTTTTCTTACAAATATTTGGAAACATTTAAAT.................. 1302
           |||   ::::::   :::||||||:::
  539 .....LeuAspThrValLeuGlnHisLeuGlyGluLysProProAlaTrp 553

 1303 ...CCTGAAGTA..............CAAACTCCAAAAGACCTTACTAT 1334
         |||::::::               :::   |||::::::::::|||::
  554 IleProAspThrTyrTyrIleGlySerAsnIleProSerAsnIleThrVa 570

 1335 AACTGAA............................AACCCTGAAGACT 1354
      :   |||                          :::|||||||||
  570 lProGluProThrProLeuProProSerAsnGluThrThrProGluAspA 587

 1355 GTAGCACAATA..................................... 1365
        :::|||
  587 snGlnThrAsnProAsnProSerGlnGlyAsnGluThrAsnProAsnPro 603

 1366 .........GATGAATATGTTCCA......GGACTCGTTATTCCAACCAT 1400
              :::|||   :::|||      :::   :::   |||:::
  604 SerProGlyAsnGluThrThrProSerAspAsnGlnThrThrProSerTh 620

 1401 AGGGATATATTACAAACACAATGAT..............GCTATATGGA 1435
       |||   :::   |||   ::::::                :::   |||
  620 rGlyAspPheValLysProGlySerLeuSerValLysValThrAspTrpG 637

 1436 AAACTAGATCTTATTCAATTCATGCACCTGGTGTAGACAGATATGAATGG 1485
        :::   :::|||:::::                      |||:::|||
  637 lyAsnThrGluTyrAspValThrLeuAsnLeuGlyGlyThrTyrAspTrp 653

 1486 GACTTGGTCAAAGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCA 1535
        :::||||||||||         |||:::   ||||||:::::::::
  654 ...ValValLysVal.........LysLeuLysAspGlySerSerValSe 666

 1536 TTATTATAACACTGACTATAAA......GAAAGCTCTATTATTATATGGA 1579
         :::   :::      |||      |||::::::   :::::::::
  666 rSerPheTrpSerAlaAsnLysAlaGluGluGlyGlyTyrValValPheT 683

 1580 AAGGGGAACCATATTTAATTAAATGGTGGCAACAAGGACCT 1620
                               |||:::::::||||||
  683 hrProValSer..............TrpAsnArgGlyPro 691

seq_name: sp_bacteria:Q9ALZ0
```

```
seq_documentation_block:
ID   Q9ALZ0      PRELIMINARY;      PRT;   563 AA.
AC   Q9ALZ0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ENDOCHITINASE A.
GN   CHIA.
OS   Serratia liquefaciens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
OX   NCBI_TaxID=614;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GM1403;
RA   Choi Y.J., Kang S.O., Ha K.J., Shin Y.C.;
RT   "Characterization of Serratia liquefaciens GM1403 endochitinase gene
RT   (chiA).";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF334683; AAK07482.1; -.
SQ   SEQUENCE   563 AA;  61117 MW;  5773C64B787BAF17 CRC64;


alignment_scores:
            Quality:  147.50              Length:    424
              Ratio:  0.749                 Gaps:     20
 Percent Similarity:  46.462   Percent Identity: 19.811

alignment_block:
 US-09-579-383-2 x Q9ALZ0   ..

 Align seg 1/1  to: Q9ALZ0  from: 1  to: 563

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   ::::::         |||:::::::::::|||::::::   :
      27 yLysProThrLeuAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||           ::::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         ||||||      :::      |||:::::::::   :::   ||||||
      57 LysSerAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspTh 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   ::::::::::::      |||:::|||
      73 rGlyThrThrAlaLysValLeuLeuAsnGlyLysGluValTrpSerGlyA 90

     247 ............................AATATAAATAAAAATGAT 264
                                     ::::::||||||:::
      90 laSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA.......................................... 267
         |||
     107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

     267 .......................................... 267

     123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuAlaP 140

     268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
              |||:::|||      :::|||||||:::   |||      ||||||
     140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

     313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
                  ::::::::::::|||:::   :::|||:::      ||
     155 AspSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171
```

```
     354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
         |         ::      :::|||:::   |||      ::::::   :
     171 y......ArgAsnPheThrValAspLysLeuProAlaGlnAsnLeuThrH 186

     404 ATATTGCATTTGCTCGCATT.........................AAT 426
         :::::   ::::::   |||                         :::
     186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAspGlyIleAsnAsp 202

     427 ATGTTATATGATGTATCTAGACCATTT.....................AA 455
            |||   :::::::::      |||                    ::
     203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

     456 TGGAAGACAAAGATTCCTATTAAGAAAACAC................... 486
         :|||||||:::   |||   :::      |||
     219 nGlyArgGluAspPheLysValSerIleHisAspProPheAlaAlaLeuG 236

     487 ..............GGCTTAGAATAT......GAAACCTATGGTATGATG 516
                       |||:::             :::   |||
     236 lnLysGlyGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

     517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
         ::::::::::::      ::::::::::::|||||||||:::   |||||
     253 PheGlyGlnLeuMetAlaLeuLysGlnAlaArgProAspLeuLysIleLe 269

     567 TTTATCCTTAGGTGGAGAAACC..................TATATGATAG 598
         |   |||:::||||||   |||                  :::|||   |
     269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

     599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
         ||      |||:::      :::|||   ::::::   ::::::::::::
     286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

     649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
            :::      :::|||||||||||||||||||||||||   |||   ||
     303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

     693 GAAGTTTTACAAC......TTAAATGAATTAAATTTTTCAAATTATTATA 736
         |:::      |||      |||:::::::   :::   :::      |||:
     319 yGlnGlyAlaAsnProLysLeuGlySerThrGlnAspGlyAlaThrTyrV 336

     737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
         :::::|||::::::   ||||||      :::      :::::::|||
     336 alGlnLeuMetLysGluLeuArgAlaMetLeu......AspGlnLeuSer 350

     787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836
         :::   :::|||   :::             |||:::::::|||:::
     351 AlaGluThrGlyArgLysTyrGluLeuThrSerAlaIleSerAla..... 365

     837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTGTCTG 886
                |||||::                                :
     366 .......GlyLysAspLys.............................IleA 371

     887 AACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCA 936
         ::::::::::
     371 spLysValAsp..................................... 374

     937 GCAGGAACTTTTATTAATATTTTTAATACAGCAAAGGAGAAAATAGATCT 986
                           :::|||||||||:::::::::::::|||
     375 ....................TyrAsnThrAlaGlnAsnSerMetAspHi 384

     987 TGTATTTATTCAAACATACAAT 1008
          :::|||::::::::::|||:::
     384 sIlePheLeuMetSerTyrAsp 391

seq_name: sp_bacteria:O83008

seq_documentation_block:
ID   O83008      PRELIMINARY;      PRT;   563 AA.
AC   O83008;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CHITINASE A PRECURSOR.
GN   CHIA.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
OX   NCBI_TaxID=615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2170;
RA   Suzuki K., Nikaidou N., Watanabe T.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2170;
RX   MEDLINE=98037511; PubMed=9371460;
RA   Watanabe T., Kimura K., Sumiya T., Nikaidou N., Suzuki K., Suzuki M.,
RA   Taiyoji M., Ferrer S., Regue M.;
RT   "Genetic analysis of the chitinase system of Serratia marcescens
RT   2170.";
RL   J. Bacteriol. 179:7111-7117(1997).
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; AB015996; BAA31567.1; -.
DR   HSSP; P07254; 1CTN.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR002173; PfkB.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
KW   Glycosidase; Hydrolase; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    563       CHITINASE A.
SQ   SEQUENCE   563 AA;  61111 MW;  CF0B130905AF6E8B CRC64;


alignment_scores:
            Quality:  144.50              Length:    426
              Ratio:  0.730                 Gaps:     19
 Percent Similarity:  46.479   Percent Identity: 19.718

alignment_block:
 US-09-579-383-2 x O83008   ..

 Align seg 1/1  to: O83008  from: 1  to: 563

      25 TTAATTATAGTATCCATCTTGTATTCTGCAAATTCCAGAACCTTGAAAGG 74
         |||:::|||   |||   |||   ||||||   :::   :::      ||
      11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGl 27

      75 AAAAAATAATATA.........AATAATTCATTGGGAATAATACGGGAAA 115
         ||||   :::|||         |||:::::::::::::|||::::::   :
      27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44

     116 ATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATCTT 165
         :::::      |||             :::::::::::::      :::
      44 spGlnAlaAlaThr............AlaTyrAsnAsnLeuValLysVal 56

     166 AAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT.. 213
         |||:::      :::      |||::::::::::   :::   ||||||
      57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlyAspTh 73

     214 .GGGTGTAACTCTAGAATTACAAAAAATAATAAA.............. 246
          |||   ::::::::::::      |||:::|||
      73 rGlyThrThrAlaLysValLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90

     247 ............................AATATAAATAAAAATGAT 264
                                        ::::::||||||:::
      90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106

     265 AGA..................................... 267
```

```
     |||
107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123

267 ..................................................  267

123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuAlaP 140

268 .....AAATCACCAAGACAAATTTTAGAGGAGTATAAAAAAAGGAAACAA 312
         |||:::|||      :::|||||||:::   |||      ||||||
140 roLeuLysGluPro......LeuLeuGluLysAsnLysProTyrLysGln 154

313 GGT.........ATTATAGCAGGATACTATGGTTCATGGAACAGTCAAGG 353
    :::         ::::::::::::|||:::   :::|||:::      ||
155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGl 171

354 TGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTTAT 403
    |       :::     :::|||:::   |||     ::::::   :
171 y......ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186

404 ATATTGCATTTGCTCGCATT..............AATATGTTATATGAT 438
    :::::   ::::::  |||               |||   :::   |||
186 isLeuLeuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202

439 ................................GTATCTAGACCATTTAA 455
                                    ::::::|||      ::
203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysGl 219

456 TGGAAGACAAAGATTCCTATTAAGAAAACAC............GGCTTAG 493
    :||||||::   |||   :::      |||           :::|||:
219 nGlyArgGluAspPheLysValSerIleHisAspProPheAlaAlaLeuG 236

494 AATATGAAACCTATGGTATGATG.......................... 516
    ::          |||:::
236 lnLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252

517 CTTAATGAAATTAGACGTATCAGAAAAGTACGTCCAGATGTAATTATTCT 566
    ::::::::::::      ::::::::::::::|||||||:::   |||||
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269

567 TTTATCCTTAGGTGGAGAAACC.................TATATGATAG 598
    |   |||:::||||||   |||                  :::|||   |
269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePhePheMetGlyA 286

599 ATATAGAAAAAGAAATTGATTATGTGGATAAAATATTGAAGCTTGTTAAT 648
    ||      |||:::      :::|||   ::::::  ::::::::::::
286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302

649 GATTTTGAT...TTAGATGGTGTAGATATTGACTGGGAA...CCACATGG 692
       :::      :::|||||||||||||||||||||||||   |||   ||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyGl 319

693 GAAGTTTTAC......AACTTAAATGAATTAAATTTTTCAAATTATTATA 736
    ||||            ||||||::::::   :::   ::::::   |||:
319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTyrV 336

737 TTAAATTAATTAACTTGTTAAGAAAAACTATTCCGGAAGAAAAGTTAATT 786
    ::   |||::::::   ||||||      :::   :::
336 alLeuLeuMetLysGluLeuArgAlaMetLeuAspGln............. 348

787 TCAATTTCTGGTTCATCAAATGCTGCATTATCATGCGTTTCAGGAGTTGC 836

348 ..................................................  348

837 ATCTTTCTGTAAAGATGAAGAATCTCCATATAACACTAAATTTTTGTCTG 886
                                        ||||||
349 ........................................LeuSer.  350

887 AACAAATAGAAACAAATAAAGAATTACATAGGGCAGCAGCGATGTTATCA 936
        :::||||||:::::::::   :::   ::::::|||:::   :::
351 .....ValGluThrGlyArgLysTyrGluLeuThrSerAlaIleSerAla 365

937 GCAGGAACTTTTATTAAT......ATTTTTAATACAGCAAAGGAGAAAAT 980
    :::         |||:::         :::|||:::|||:::::::::::
```

```
366 GlyLysAspLysIleAspLysValAlaTyrAsnValAlaGlnAsnSerMe 382

981 AGATCTTGTATTTATTCAAACATACAAT 1008
    :|||   :::|||:::::::::|||:::
382 tAspHisIlePheLeuMetSerTyrAsp 391
```

seq_name: sp_invertebrate:O96206

seq_documentation_block:
ID O96206 PRELIMINARY; PRT; 1817 AA.
AC O96206;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PREDICTED SECRETED PROTEIN.
GN PFB0565W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001403; AAC71902.1; -.
SQ SEQUENCE 1817 AA; 217996 MW; D1812785960E0BCE CRC64;

alignment_scores:
Quality: 144.00 Length: 558
Ratio: 0.514 Gaps: 31
Percent Similarity: 50.179 Percent Identity: 20.968

alignment_block:
US-09-579-383-2 x O96206 ..

Align seg 1/1 to: O96206 from: 1 to: 1817

```
 64 ACCTTGAAAGGAAAAAATAATATAAATAATTCATTGGGAATAATACGGGA 113
    |||   :::||||||:::::::::|||:::
355 ThrLysGlnGlyLysHisAspMetAsnAsp.................Ty 365

114 AAATAAAAATAAAACTCATCAAACGGAAATACATGAGTCTTTTTCACATC 163
     |||   ||||||   ::::::   :::|||:::   ::::::::::::
365 rAsnMetAsnLysLysAsnAsnMetAspIleAsnIleThrIleAsnAsnA 382

164 TTAAATCGAATAATAGTAATTTTGTAGAATATGGATCTTATTGCGGAGAT 213
      ::::::||||||:::|||   :::            |||      :::
382 snAsnAsnAsnAsnAsnLysIle............TyrAsnAspAsn 394

214 GGGTGTAACTCTAGAATTACAAAAAATAATAAAAATATAAATAAA..... 258
          |||      :::   :::|||:::   ||||||::::::
395 ThrLeuAsn......ValTyrAsnAsnSerTyrAsnIleHisSerAsnHi 409

259 .......AATGATAGAAAATCACCAAGACAAATTTTAGAGGAGTATAAAA 301
          ||||||:::::::::      |||:::||||||::::::   |
409 sLeuMetAsnAspLysArgLysAsnAlaGlnValLeuGluLysHisLeuL 426

302 AAAGGAAACAAGGTATTATAGCAGGATACTATGGTTCATGGAACAGTCAA 351
    ||            :::::::::      :::      :::|||   :::
426 ys............MetLeuCysAspAsnPhe......PheAsnLeuGlu 436

352 GGTGATAGAGCAAAACATATGATTGATTCAAACCCAATGGTGTCAATTTT 401
              ::::::::::::|||   :::|||      ::::::
437 GluPheTyrSerSerAsnIleIleIleAsnAsnMetAspIleGluTyrTh 453

402 ATATATTGCATTTGCTCGCATTAATATGTTATATGATGTATCTAGACCAT 451
```

```
 453 rTyrAspTyrPhe.....IleLeuTyrGlu.....LysCysP 464

 452 TTAATGGAAGACAAAGATTCCTATTAAGAAAACACGGCTTAGAATATGAA 501
 464 heLeuProIleGluArgIleVal.....HisValAsnTyrMet 476

 502 ACCTATGGTATGATGCTTAATGAAATTAGACGTATCAGAAAAGTACGTCC 551
 477 LysTyr...LeuTyrLysAsnAsnGluArgLysLysAsnLysIleArgLy 492

 552 AGATGTAATTATTCTTTTATCCTTAGGTGGAGAAACCTATATGATAGATA 601
 492 sPheLeuIleThrLeuLeuGlu.....TyrSerArgAspI 504

 602 TAGAAAAAGAAATTGATTATGTGGATAAAATATTG...AAGCTTGTTAAT 648
 504 leAsnPheHisPhePheIlePheAsnLeuIleLeuTyrLysCysLysAsn 520

 649 GATTTTGATTTAGATGGTGTAGATATTGACTGGGAACCACATGGGAAGTT 698
 521 GluPheProCysSerIlePheGluLeuHisIleSerGlnTyrLeuTyrPh 537

 699 TTAC...AACTTAAATGAATTAAATTTTTCAAATTATTATATT..... 738
 537 ePheValLysLeuAsnGluLeuAsnIleLysAspAlaTyrIleTyrTyrP 554

 739 .....AAATTAATTAACTTGTTAAGAAAAACT 765
 554 heAsnAsnPheLysTyrGlnAspMetIleIleTyrPheSerArgLysAla 570

 766 ATT...CCGGAAGAAAAGTTAATTTCAATTTCTGGTTCATCAAATGCTGC 812
 571 PheTyrProTrpGlu.....ThrAsnValGluGlnGlnLysLysGlnTh 585

 813 ATTATCATGCGTTTCAGGAGTTGCATCTTTCTGTAAAGATGAAGAATCTC 862
 585 rLeuSerTyrIle..... 589

 863 CATATAACACTAAATTTTTGTCTGAACAAATAGAAACAAATAAAGAATTA 912
 590 ..TyrAsnAspLys.....IleLysLysAsnLysLys... 599

 913 CATAGGGCAGCAGCGATGTTATCAGCAGGAACTTTTATTAAT..... 954
 600 AsnAsnSerTyrTyrGluMetAsnAsnAsnThrTyrMetAsnGluHisGl 616

 955 .....ATTTTTAATACAGCAAAGGAGAAAATAG 982
 616 yTyrThrAspIleGluAsnGluArgLeuAsnLysLysAsnLysArgLeuA 633

 983 ATCTTGTATTTATTCAAACATACAATTTAGAAACTACAAATCCAGATATA 1032
 633 snVal.....ArgGlyArgThrAsnThrLeuAspAspIle 644

1033 ATGGTAGATATGTACTTATCCCATTTATATTTTGGTTTAAAATATAACAT 1082
 645 IleValSer.....AspHisGlyAsnSerTyrAspLysTyrAsnTh 658

1083 CACAATCATATTAGGTTTTTCATTAGAACATAACAGAGGTGGATTTAGTC 1132
 658 rSer.....LysHisAsnArg..... 663

1133 CCGAAAATAAAGAATTATTAGAATTGGTAGGAAAAACAATACATGATAAA 1182
 664 .....ArgLysAsnHisIleAsnGluMetLysLysLysGlnAsnAsnLys 678

1183 AATCAAAATAATAATAGGGCAGATGGTATAGGGATATGGCATTTATTTAT 1232
 679 LysLysAsnThrLeuPheValAspGlyLysAspMetGluGlyIleGlyLy 695

1233 GAAAGAACAATTACCAACTGGATCATTCGATGTAGATATTTTTCTTACAA 1282
```

```
 695 sGluLysGluLysGluAsnLysAsnMetAsnAsnAsnIlePheTyrAsnA 712

1283 AT.....ATTTGGAAACATTTAAATCCTGAA 1308
 712 snSerTyrSerAsnIleAsnAsnSerSerTyrSerAsnIleAsnAsnAsp 728

1309 GTACAAACTCCAAAAGACCTTACTATAACTGAAAACCCTGAAGACTGTAG 1358
 729 IleTyrSerValAspAsnMetThrSerValAsnAsnThrLysTyrValSe 745

1359 CACAATAGATGAATATGTTCCAGGACTCGTTATTCCAACCATAGGGATAT 1408
 745 rGlyValProSerTyrAlaHisValLeuIleAsnLysGlnValAsnGluT 762

1409 ATTACAAA.....CACAATGATGCTATATGGAAAACTAGATCT 1446
 762 yrTyrGlnGlyLeuProAsnTyrAsnAsnMetMetIleLys...GlySer 777

1447 TATTCAATTCATGCACCTGGTGTAGACAGATATGAATGGGACTTGGTCAA 1496
 778 HisIleIleAsnGluLeuProLysAsnAsnTyr..... 788

1497 AGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCCATTAT..... 1539
 789 ....IleTyrGluAsnAsnTyrIleGlyGln.....AsnTyrLeuMetT 802

1540 .....TATAACACTGACTATAAAGAAAGCTCTATTATTATATGG 1578
 802 hrAsnProLeuTyrAsnLysGluThrLysAspIlePheTyrThrIleTyr 818

1579 AAAGGGGAACCATATTTAATTAAA 1602
 819 Lys.....TyrLeuPheLys 823
```